Modélisation des séries temporelles

Séance 5

Modélisation avec variable exogène:
modèles à « fonction de transfert »

Frédéric Sur

https://members.loria.fr/FSur/enseignement/modseries/

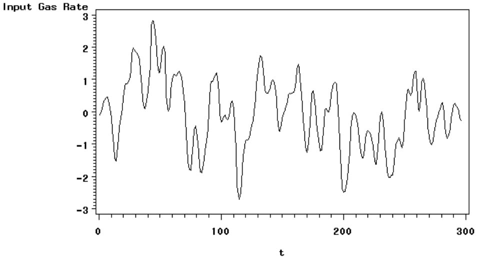
Séance 5

Exemples

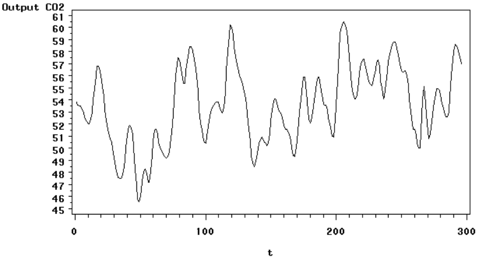
- modèles proie-prédateur : nombre de proies lié au nombre de prédateurs, avec effet retard pour les naissances
- construction de logements neufs en fonction de prêts immobiliers accordés, effet retard
- émission de CO_2 d'une chaudière fonction du débit de gaz en entrée $\rightarrow$ voir exemple
- etc.

Exemple de la chronique CO2 (series J de Box&Jenkins)

a) Debit alimentation en gaz : input x



b) Concentration en CO2 : output y



Exemple de la chronique CO2 (cf poly)

X_t : débit de gaz en entrée d'une chaudière
Y_t : émission de CO2

Modèle identifié :

$$Y_t = 53.26 \; - \; \frac{0.535 + 0.376B + 0.518B^2}{1 - 0.548B} X_{t-3} + \frac{1}{1 - 1.532B + 0.632B^2} \varepsilon_t \qquad (1)$$

Rappel : $\frac{1}{1-0.548B} = 1 + 0.548B + 0.548^2 B^2 + \dots$
donc $\frac{0.535+0.376B+0.518B^2}{1-0.548B} =$
$0.535 + (0.367 + 0.535 \cdot 0.548)B + (0.518 + 0.376 \cdot 0.548 + 0.535 \cdot 0.548^2)B^2 + \dots$
et $B^k X_{t-3} = X_{t-3-k}$

Intérêt double :
$\longrightarrow$ modélisation
$\longrightarrow$ prévision

Modèles à variable exogène

Modèle à fonction de transfert :

cf régression dynamique

cf ARIMAX : ARIMA with exogeneous factors

$$\begin{aligned} Y_t &= c - \alpha_1 Y_{t-1} - \cdots - \alpha_l Y_{t-l} \\ &\quad + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_k X_{t-k} + u_t \end{aligned} \tag{2}$$

avec :

- (Y_t) est la **chronique à modéliser**
- (X_t) est la **chronique explicative**
- (u_t) est la **chronique des erreurs**

$\rightarrow$ les chroniques sont supposées **stationnaires**

$\rightarrow$ $(1 + \alpha_1 B + \cdots + \alpha_l B^l)(Y_t - \mu) = (\beta_0 + \cdots + \beta_k B^k)X_t + u_t$
d'où la représentation avec les fractions rationnelles en B

$\rightarrow$ éventuellement $\beta_i = 0$ pour $i < b$ (b est le **retard**),

Écriture avec l'opérateur de décalage B

Modèle à fonction de transfert :

$$Y_t = \mu + \frac{\Omega(B)}{\Delta(B)} X_{t-b} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \tag{3}$$

avec :

- $b \geqslant 0$ le retard
- $\Omega(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$
- $\Delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$
- $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$
- $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$

et (ε_t) bruit blanc gaussien d'écart-type σ

Définition : $\frac{\Omega(B)}{\Delta(B)}$ est appelé *fonction de transfert*.

Rappel : $\frac{\Theta(B)}{\Phi(B)}\varepsilon_t$ est la représentation ARMA d'un processus stationnaire linéaire gaussien.

Séance 5

« Stationnarisation »

Dans le modèle, X_t et Y_t doivent être *stationnaires*...

$\longrightarrow$ si ce n'est pas le cas, il faut éventuellement dériver au préalable les chroniques étudiées (de la même manière) pour que le résultat soit stationnaire.

(voir la discussion sur les régressions fallacieuses)

Blanchiment de la chronique explicative

X_t : chronique explicative

Pré-blanchiment (de X_t)
(X_t) est un processus stationnaire (« donc » ARMA).

→ il existe des polynômes Φ_1 et Θ_1 t.q. :

$$X_t = (\mu +) \frac{\Theta_1(B)}{\Phi_1(B)} \chi_t \tag{4}$$

où χ_t est un bruit blanc gaussien.

On écrit

$$\frac{\Phi_1(B)}{\Theta_1(B)} X_t = \chi_t \tag{5}$$

→ le filtre $\frac{\Phi_1(B)}{\Theta_1(B)}$ « blanchit » la chronique X_t.

Blanchiment de la chronique explicative

$$Y_t = (\mu +) \, \frac{\Omega(B)}{\Delta(B)} X_{t-b} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \tag{6}$$

On applique $\frac{\Phi_1(B)}{\Theta_1(B)}$ à Y_t :

$$\Upsilon_t \; = \; \frac{\Phi_1(B)}{\Theta_1(B)} Y_t \; = \; \frac{\Omega(B)}{\Delta(B)} \chi_{t-b} + \widetilde{\varepsilon}_t \tag{7}$$

(les polynômes symboliques commutent)

avec :

- $\widetilde{\varepsilon}_t = \frac{\Phi_1(B)\Theta(B)}{\Theta_1(B)\Phi(B)} \varepsilon_t$ est *stationnaire* (pas un bruit blanc comme ε_t), supposé *indépendant* de χ_t (sinon problème d'identification. . .)
- et χ_t b.b.g. par définition de Φ_1/Θ_1.

Le corrélogramme croisé

On est ramené à l'identification du modèle :

$$\Upsilon_t = \frac{\Omega(B)}{\Delta(B)} \chi_{t-b} + \widetilde{\varepsilon}_t \tag{8}$$

Comment identifier Ω et Δ ? (ou la fonction de transfert Ω/Δ)

$\rightarrow$ nous allons utiliser l'outil suivant :

Définition : le *corrélogramme croisé*

$$\rho(h) = \frac{\mathrm{Cov}(\chi_t, \Upsilon_{t+h})}{\sqrt{\mathrm{Var}(\chi_t) \cdot \mathrm{Var}(\Upsilon_t)}}$$

(ne dépend pas de t, cf slide suivant)

Remarque : $\rho(h)$ est défini pour $h \in \mathbb{Z}$
(a priori, $\rho(h) \neq \rho(-h)$)

Identification de la fonction de transfert

$$\Upsilon_t = \frac{\Omega(B)}{\Delta(B)}\chi_{t-b} + \widetilde{\varepsilon}_t \quad \text{où} \ \ (\chi_t) \text{ b.b.g. et} \quad (\widetilde{\varepsilon}_t) \text{ stationnaire}$$

En inversant Δ, on écrit : $\Upsilon_t = \sum_{h \geqslant 0} \nu_h \chi_{t-h} + \widetilde{\varepsilon}_t$
(ici si $h < b, \nu_h = 0$)

Proposition *(intérêt du corrélogramme croisé)*

$$\forall h \geqslant 0,\ \rho(h) \ = \ \nu_h \frac{\sigma_\chi}{\sigma_\Upsilon} \qquad \text{et} \ \ \forall h < 0,\ \rho(h) = 0$$

Preuve : (basée sur hypothèses sur χ et Υ !)

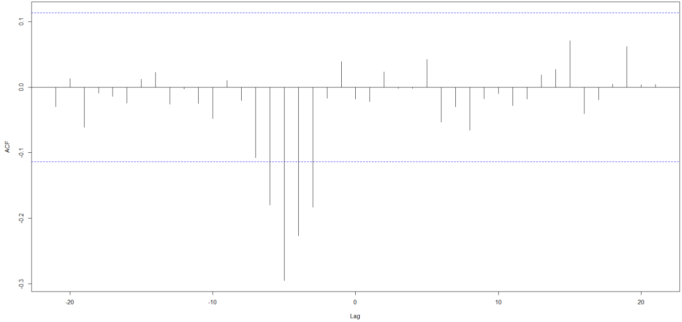
$$\mathrm{Cov}(\chi_t, \Upsilon_{t+h}) = \mathrm{Cov}\Big(\chi_t, \sum_{h' \geqslant 0} \nu_{h'} \chi_{t+h-h'}\Big) \quad (\text{car } \chi_t \text{ et } \widetilde{\varepsilon}_t \text{ décorrélés})$$

$$= \begin{cases} \nu_h \,\mathrm{Var}(\chi) & \text{si } h \geq 0 \\ 0 & \text{sinon} \end{cases} \qquad (\text{car } \chi_t \text{ bruit blanc})$$

Conclusion : le coef. ν_h du "polynôme" Ω/Δ est **proportionnel** au coef. $\rho(h)$ du corrélogramme croisé entre χ et Υ.

Exemple de corrélogramme croisé : `ccf`

(corrélogramme croisé entre χ_t et Υ_t, pas entre X_t et Y_t !)



Attention : convention R vs. autres logiciels et définition...

`help(ccf)`

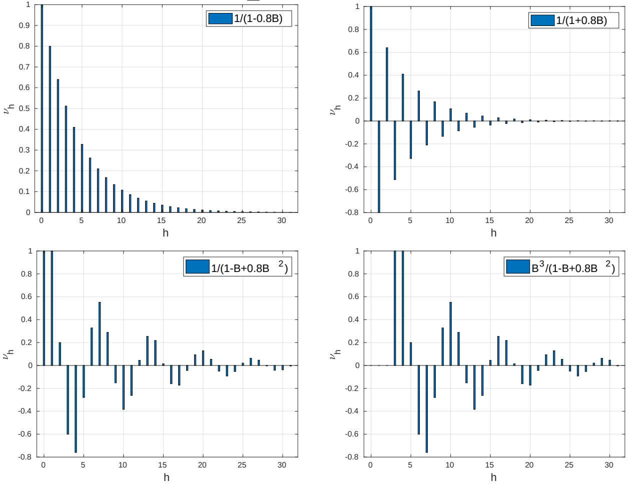
"the lag k value returned by `ccf(x,y)` estimates the correlation between `x[t+k]` and `y[t]`"

→ donc (par stationnarité) : corrélation entre `x[t]` (χ) et `y[t-k]` (Υ)

→ `ccf(x,y)` correspond donc au graphe de $\rho(-h)$: on doit observer des corrélations significatives pour des « lags » négatifs

Allures des corrélogrammes selon la fonction de transfert

Rappel : ν_h t.q. $\dfrac{\Omega(B)}{\Delta(B)} = \sum_{h\geq 0} \nu_h B^h$ & $\rho(h) = \nu_h \, \sigma_\chi / \sigma_\Upsilon$



et bien entendu, $B^3 + 0.2B^4 - 0.1B^5$ donne un corrélogramme avec des « pics » en $h = 3, 4, 5$, de hauteur proportionnelle à 1, 0.2, -0.1

Pratique du corrélogramme croisé

$$Y_t = \frac{\Omega(B)}{\Delta(B)} X_{t-b} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t = \nu'(B) X_{t-b} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \qquad (9)$$

1. premier pic (significativement) non nul sur le corrélogramme donne le décalage b,
2. si décroissance (assez) lente du corrélogramme : on envisage un dénominateur (cf inversion de $(1 - \lambda B)$)
   si sinusoïde : dénominateur de degré $\geqslant 2$,
   (toujours avec l'objectif d'un *modèle simple*)
3. sinon le nombre (et la position) des pics non nuls donne le degré du numérateur (et les coefficients non nuls).

Remarque : si le corrélogramme croisé a des pics significatifs pour des décalages « futurs », le modèle n'est pas bien adapté (non-causalité)

Identification du modèle : ARMA sur les résidus

On cherche un modèle du type :

$$Y_t = \frac{\Omega(B)}{\Delta(B)} X_{t-b} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \tag{10}$$

Avec l'étape précédente, on connaît les degrés de Ω et Δ (et éventuellement la position des coef. nuls), et le décalage b.

Ensuite :

Calcul de la fonction de transfert

On revient aux séries (stationnaires) initiales :

$$Y_t = \frac{\Omega(B)}{\Delta(B)} X_{t-b} + u_t \tag{11}$$

$\rightarrow$ on estime les coefficients de Ω et Δ, résidus u_t

$\rightarrow$ ACF / PACF des résidus u_t : ils permettent d'identifier Θ et Φ

Modèle ARMA sur les résidus stationnaires u_t

$\rightarrow$ Estimation de Θ et Φ et réestimation de Ω et Δ.

Résumé : identification et estimation

$$Y_t = \mu + \frac{\Omega(B)}{\Delta(B)} X_{t-b} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \tag{12}$$

1. Pré-blanchiment de X_t.

2. Corrélogramme croisé sur les séries préfiltrées.
(on utilise le filtre blanchissant X_t pour X_t et Y_t)
Permet l'identification du retard et de la forme de la fonction de transfert.

3. Estimation de la fonction de transfert $\Omega(B)/\Delta(B)$
et identification/estimation du modèle ARMA $\Theta(B)/\Phi(B) \cdot \varepsilon_t$ des résidus u_t.

Remarque : bien sûr, vérifications habituelles de rigueur
(résidus ε_t du modèle final = bruit blanc gaussien, paramètres significatifs, etc).

Remarque : régression fallacieuse

On a supposé les chroniques X_t et Y_t *stationnaires*.
$\rightarrow$ et si ce n'est pas le cas ?

Exemple : soient X_t et Y_t intégrées d'ordre 1
(X_t et Y_t non stationnaires, $(1-B)X_t$ et $(1-B)Y_t$ stationnaires)

Alors $Y_t - aX_t - b = R_t$ est généralement aussi intégrée d'ordre 1.

$\rightarrow$ si estimation de a, b par régression (moindres carrés des R_t) :
les t-tests (Student) sont trop "optimistes"...
(car R_t n'est pas un b.b.g., mais une marche aléatoire ou un ARIMA(p,1,q), donc $\text{Var}(\widehat{a}) \neq \sigma^2(X^TX)^{-1}$)

On parle de *régression fallacieuse* (*spurious regression*).

`http://tylervigen.com/spurious-correlations`

$\rightarrow$ Pour contourner le problème, on « régresse » $(1-B)Y_t$ sur $(1-B)X_t$, avec des résidus ARMA (et pas i.i.d.)...

Séance 5

Modèles à variable exogène

Modèle à variable exogène et fonction de transfert :

$$Y_t = \mu + \frac{\Omega(B)}{\Delta(B)} X_{t-b} + u_t \tag{13}$$

Rappel : Modèle d'*intervention* :

$$X_t = \mu + \alpha I_t^{t_0} + u_t = \mu + \frac{\Omega(B)}{\Delta(B)} I_t^{t_0} + u_t \tag{14}$$

$\rightarrow$ modèle similaire : $\qquad I_t^{t_0} \longleftrightarrow X_t$

avec $I_{t_0}(t)$ déterministe / (X_t) stochastique stationnaire.

Donc fonctions R similaires, mais pas d'interprétation du corrélogramme croisé pour les modèles d'intervention, ni de pré-blanchiment (qui n'ont pas de sens dans ce dernier cas)

Dernières séances. . .

- **Tout retour est le bienvenu**
  (en ayant en tête les multiples contraintes. . .)
  Cf pad sur Arche

- *Lundi 23 mars 2020*
  Discussion d'un exemple, Q&R
  TP de synthèse

- *Lundi 30 mars 2020*. **Test**

  Instructions à venir

Exemple : série J de Box et Jenkins

Chronique d'entrée X_t : débit de gaz alimentant un appareil de chauffage

Chronique de sortie Y_t : concentration en CO2

$\rightarrow$ cf carnet Jupyter et code sur page web