Désaisonnaliser une série temporelle

6 - Le modèle Reg-ARIMA

Alain Quartier-la-Tente

Objectifs de cette séquence

Objectifs : modélisation Reg-ARIMA, pré-ajustement de X13-ARIMA.

Après cette séquence, vous saurez :

- La structure et les fonctions d'un modèle Reg-ARIMA
- Reconnaître les modèles de JD+ à partir des diagnostics
- Modifier les spécifications du modèle

Questions de positionnement

Qu'est-ce qu'un processus stationnaire ?

Tendance, cycle, saisonnalité sont-ils des processus stationnaires ?

Que signifie "ARIMA" et que reflète un tel modèle ?

Comment se comportent les erreurs de prévision d'un modèle ARIMA ?

Qu'est-ce qu'un SARMA ?

Saurons nous "deviner" le comportement de la saisonnalité à travers un modèle ARIMA ?

Que sont les critères d'information et à quoi ça sert ?

X13-ARIMA

Deux modules :

- Reg-ARIMA : phase de pré-ajustement
  - Régression linéaire pour correction préalable des « non-linéarités »
  - Modélisation ARIMA pour faire des prévisions
  - Deux étapes indépendantes en schéma, mais traitements itératifs !
- X11 : phase de décomposition

La partie « régression linéaire » de X13-ARIMA

Objectif : supprimer les « non-linéarités » par régression linéaire :

- outliers
- effets de calendrier
- autres régresseurs éventuels (ex : température)

$$Y_t = \sum \hat{\alpha}_i O_{it} + \sum \hat{\beta}_j C_{jt} + X_t$$

Série *linéarisée* : $X_t = Y_t - \sum \hat{\alpha}_i O_{it} - \sum \hat{\beta}_j C_{jt}$

GROS résidu de la régression

N'est pas le résidu Reg-ARIMA, qui est un bruit blanc

La décomposition est réalisée sur la série linéarisée

Sommaire

Quelques définitions (1/2)

Série temporelle : suite de variables aléatoires $(X_t)_t$ dont on observe une réalisation $(X_t(\omega))_t$

La suite $(X_t)_t$ est appelée *processus stochastique*

Quelques définitions (1/2)

Série temporelle : suite de variables aléatoires $(X_t)_t$ dont on observe une réalisation $(X_t(\omega))_t$

La suite $(X_t)_t$ est appelée *processus stochastique*

Un processus est dit *stationnaire* lorsque la loi de X_t n'évolue pas dans le temps : distribution $\forall s$, $(X_t, \dots, X_{t+s})$ indépendante du temps

$\implies$ série plus ou moins horizontale et de variance constante

➔ Notion pour faire l'inférence et construire un modèle ARIMA

Quelques définitions (2/2)

Stationnarité, hypothèse invérifiable ➔ en pratique processus *faiblement stationnaire* :

- les moments d'ordre 2 existent
- espérance constante
- covariance entre t et $t-h$ ne dépend pas du temps, mais de la distance h $\implies$ variance constante

Quelques définitions (2/2)

Stationnarité, hypothèse invérifiable ➔ en pratique processus *faiblement stationnaire* :

- les moments d'ordre 2 existent
- espérance constante
- covariance entre t et $t-h$ ne dépend pas du temps, mais de la distance h $\implies$ variance constante

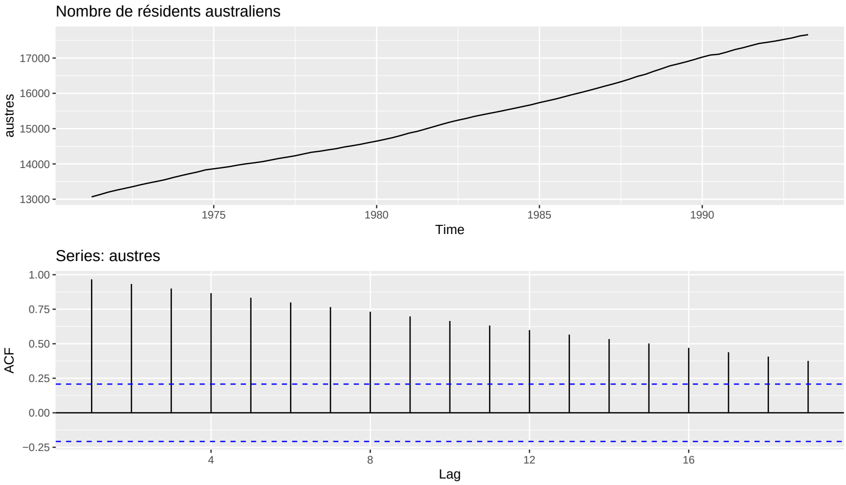
Exemple : un bruit blanc, i.e. :

- espérance nulle
- covariance entre t et $t-h$ nulle, pour tout $h \neq 0$
- variance non nulle et constante

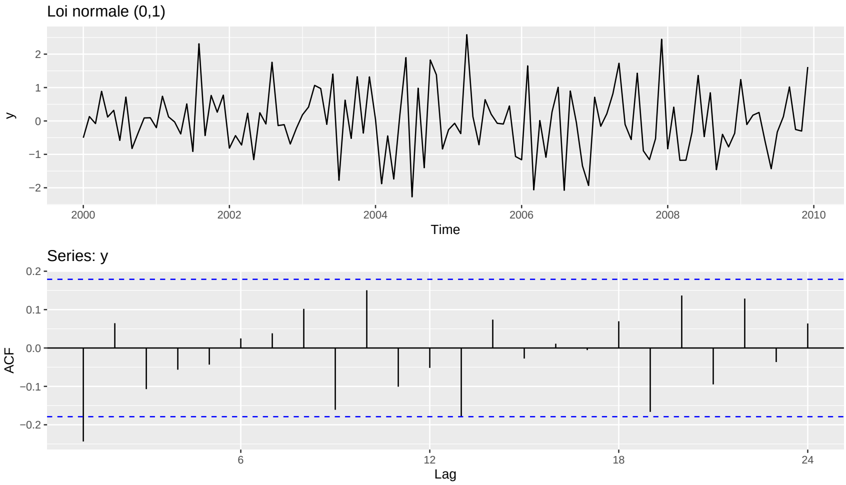
Comment identifier une série non-stationnaire (en niveau) ?

- Tracer le chronogramme
- Etudier l'ACF :
  - Série non-stationnaire : tend lentement vers 0 et $\hat{\rho}(1)$ souvent positif et élevé
  - Série stationnaire : tend rapidement vers 0

Exemple


Nombre de résidents australiens
17000
16000
15000
14000
13000
austres
1975
1980
1985
1990
Time
Series: austres
1.00
0.75
0.50
0.25
0.00
−0.25
ACF
4
8
12
16
Lag

Exemple


Loi normale (0,1)
y
2
1
0
−1
−2
2000
2002
2004
2006
2008
2010
Time
Series: y
ACF
0.2
0.1
0.0
−0.1
−0.2
6
12
18
24
Lag

La différenciation pour stabiliser le niveau

- Si la série différenciée est un bruit blanc de moyenne nulle (marche aléatoire) :

$$(I-B)y_t = y_t - y_{t-1} = \varepsilon_t \implies y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

➔ Modèle naïf
Généralement mouvement à la hausse ou à la baisse aléatoire,

La différenciation pour stabiliser le niveau

- Si la série différenciée est un bruit blanc de moyenne nulle (marche aléatoire) :

$$(I - B)y_t = y_t - y_{t-1} = \varepsilon_t \implies y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

➔ Modèle naïf
Généralement mouvement à la hausse ou à la baisse aléatoire,

- Si la série différenciée est un bruit blanc de moyenne non nulle (marche aléatoire avec dérive / *drift*) :

$$(I - B)y_t = c + \varepsilon_t \implies y_t = y_0 + ct + \sum_{i=1}^{t} \varepsilon_i$$

La différenciation pour stabiliser le niveau

- Si la série différenciée est un bruit blanc de moyenne nulle (marche aléatoire) :

$$(I-B)y_t = y_t - y_{t-1} = \varepsilon_t \implies y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

➔ Modèle naïf
Généralement mouvement à la hausse ou à la baisse aléatoire,

- Si la série différenciée est un bruit blanc de moyenne non nulle (marche aléatoire avec dérive / *drift*) :

$$(I-B)y_t = c + \varepsilon_t \implies y_t = y_0 + ct + \sum_{i=1}^{t} \varepsilon_i$$

- Parfois on a besoin de différencier plusieurs fois $(I-B)^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ ou de faire une différenciation saisonnière $(I-B^m)y_t = y_t - y_m$

La différenciation pour stabiliser le niveau

- Si la série différenciée est un bruit blanc de moyenne nulle (marche aléatoire) :

$$(I - B)y_t = y_t - y_{t-1} = \varepsilon_t \implies y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

➔ Modèle naïf
Généralement mouvement à la hausse ou à la baisse aléatoire,

- Si la série différenciée est un bruit blanc de moyenne non nulle (marche aléatoire avec dérive / *drift*) :

$$(I - B)y_t = c + \varepsilon_t \implies y_t = y_0 + ct + \sum_{i=1}^{t} \varepsilon_i$$

- Parfois on a besoin de différencier plusieurs fois $(I - B)^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ ou de faire une différenciation saisonnière $(I - B^m)y_t = y_t - y_m$
- Si saisonnalité importante, commencer par la différenciation saisonnière

Modèles Intégrés (1/3)

Soit X, processus « tendance linéaire » :

$$X_t = \alpha + \beta t + \varepsilon_t$$

Calculer l'espérance et la variance de la v.a. X_t ?
X est stationnaire ?

Modèles Intégrés (1/3)

Soit X, processus « tendance linéaire » :

$$X_t = \alpha + \beta t + \varepsilon_t$$

Calculer l'espérance et la variance de la v.a. X_t ?
X est stationnaire ?

Différence d'ordre 1 :

$$(I - B)X_t = ?$$

La série obtenue est-elle stationnaire ?

Si X est un processus « tendance polynomiale d'ordre 2 », comment stationnariser la série ?

Modèles Intégrés (2/3)

Soit X, processus « saisonnier stable » :

$$X_t = S_t + \varepsilon_t \quad \text{avec} \quad \forall t,\, S_t = S_{t+s}$$

X stationnaire ?

Modèles Intégrés (2/3)

Soit X, processus « saisonnier stable » :

$$X_t = S_t + \varepsilon_t \quad \text{avec} \quad \forall t,\ S_t = S_{t+s}$$

X stationnaire ?

Différence d'ordre 1, avec retard d'ordre s :

$$(I - B^s)X_t = ?$$

La série obtenue est-elle stationnaire ?

Si $X_t = a + bt + S_t + \varepsilon_t$, que donnerait cette différenciation ?

Modèles Intégrés (3/3)

Une différenciation « simple » d'ordre d supprime les tendances polynomiales d'ordre d :

$$(I - B)^d X_t$$

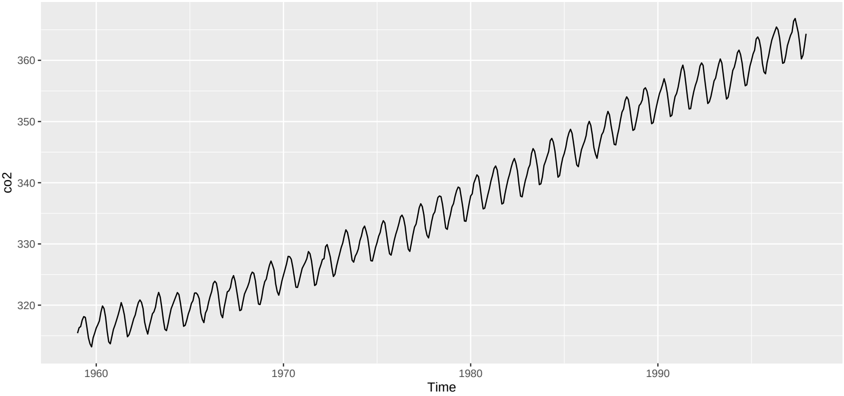
Une différenciation « saisonnière » supprime aussi les tendances linéaires :

$$(I - B^s) X_t$$

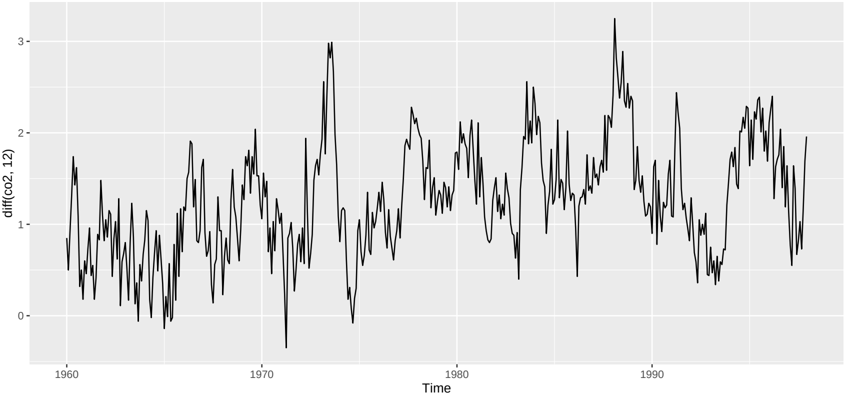
Une différenciation « saisonnière » d'ordre D plus grand que 1 est rare :

$$(I - B^s)^D X_t$$

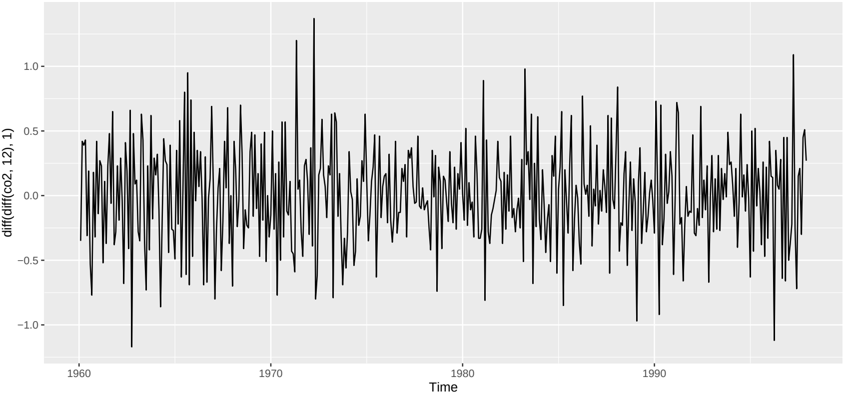
Exemple



Exemple



Exemple



Faut-il toujours différencier ?

Pour modéliser une série avec tendance on peut distinguer deux types de non-stationnarité :

1. Modèle trend-stationnaire :

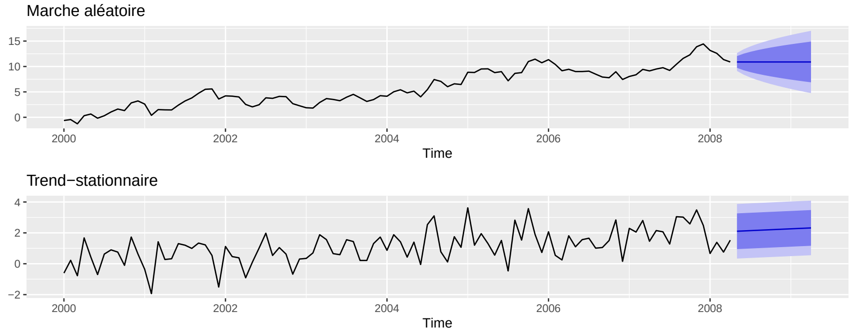
$$X_t = a + bt + \varepsilon_t$$

2. Modèle avec racine unité

$$(1-B)Y_t = b + \eta \implies Y_t = a + bt + \underbrace{\sum_{i=1}^{t} \eta_t}_{\text{tend. stochastique}}$$

On a $\mathbb{V}[X_t] = \mathbb{V}[\varepsilon_t] = cst$ indépendante du temps mais $\mathbb{V}[Y_t] = t\mathbb{V}[\eta_t]$

Exemple


Marche aléatoire
15
10
5
0
2000
2002
2004
2006
2008
Time
Trend−stationnaire
4
2
0
−2
2000
2002
2004
2006
2008
Time

Sommaire

La partie « modélisation ARIMA »

ARIMA, modèle auto-projectif :

$$X_t = f(X_{t-1}, X_{t-2}, X_{t-3}, \dots, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2} \dots)$$

Trouver f ?

Sous hypothèse de stationnarité, il existe un « modèle ARMA » qui approche la série.

Conséquence (th de Wold) : erreurs de prévision se comportent comme le résidu du modèle (bruit blanc)

On privilégie les modèles avec faible nombre de paramètres.

Méthode de Box et Jenkins pour estimer et juger de la qualité des modèles.

Sommaire

Modèles Autorégressifs (AR)

B opérateur retard : $B(X_t) = X_{t-1}$, et $B^p(X_t) = X_{t-p}$

Modèles Autorégressifs (AR)

B opérateur retard : $B(X_t) = X_{t-1}$, et $B^p(X_t) = X_{t-p}$

Modèle *autorégressif* d'ordre p, $AR(p)$:

$$
\begin{aligned}
& X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \\
\iff & (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = \varepsilon_t \\
\iff & \Phi(B) X_t = \varepsilon_t
\end{aligned}
$$

Modèles Autorégressifs (AR)

B opérateur retard : $B(X_t) = X_{t-1}$, et $B^p(X_t) = X_{t-p}$

Modèle *autorégressif* d'ordre p, $AR(p)$:

$$\begin{aligned} & X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \\ \iff & (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = \varepsilon_t \\ \iff & \Phi(B) X_t = \varepsilon_t \end{aligned}$$

ε_t *innovation* du processus (bruit blanc indépendant du passé de X)

Un AR modélise l'influence des p réalisations passées sur la réalisation courante : effet mémoire

Exemples :

- AR(1) : niveau d'un lac ;
- AR(2) : nombre de tâches solaires - Yules

Modèles « Moving Average » (MA)

Modèle *moyenne mobile* d'ordre q, $MA(q)$:

$$\begin{aligned} & X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ \iff & X_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \varepsilon_t \\ \iff & X_t = \Theta(B) \varepsilon_t \end{aligned}$$

Modèles « Moving Average » (MA)

Modèle *moyenne mobile* d'ordre q, $MA(q)$:

$$\begin{aligned} & X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ \iff & X_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \varepsilon_t \\ \iff & X_t = \Theta(B) \varepsilon_t \end{aligned}$$

Processus MA toujours stationnaire

Résulte d'une accumulation non persistante de “q” chocs indépendants

Phénomènes qui fluctuent autour d'une moyenne : MA(1) avec une constante

Exemples :

- Jeu de fléchettes

Modèles ARMA

Modèles $ARMA(p,q)$: combine $AR(p)$ et $MA(q)$, sans ou avec constante

$$\Phi(B)X_t = \Theta(B)\varepsilon_t$$

$$\Phi(B)X_t = \mu + \Theta(B)\varepsilon_t$$

Processus $ARMA$ résulte de l'effet "mémoire" et d'une accumulation non persistante de chocs aléatoires indépendants

Modèles SARMA

Modèle $SARMA(P, Q)$: $ARMA$ avec polynôme d'ordre s (4 séries trimestrielles, 12 séries mensuelles) :

$$\Phi(B^s)X_t = \Theta(B^s)\varepsilon_t \text{ ou } \Phi_s(B)X_t = \Theta_s(B)\varepsilon_t$$

Intérêt :

- montrer autocorrélations d'ordre s
- simplifier l'écriture par factorisation

Modèles SARMA

Modèle $SARMA(P, Q)$: $ARMA$ avec polynôme d'ordre s (4 séries trimestrielles, 12 séries mensuelles) :

$$\Phi(B^s)X_t = \Theta(B^s)\varepsilon_t \text{ ou } \Phi_s(B)X_t = \Theta_s(B)\varepsilon_t$$

Intérêt :

- montrer autocorrélations d'ordre s
- simplifier l'écriture par factorisation

$ARMA(p, q)(P, Q)$ combine parties régulière et saisonnière : $ARMA(p, q) \times SARMA(P, Q)$.

Identique à $ARMA(p + P * s, q + Q * s)$

Exemple série mensuelle : $ARMA(1, 1)(1, 1) = ARMA(13, 13)$

Modèles Intégrés (1/3)

Soit X, processus « tendance linéaire » :

$$X_t = \alpha + \beta t + \varepsilon_t$$

Calculer l'espérance et la variance de la v.a. X_t ?
X est stationnaire ?

Modèles Intégrés (1/3)

Soit X, processus « tendance linéaire » :

$$X_t = \alpha + \beta t + \varepsilon_t$$

Calculer l'espérance et la variance de la v.a. X_t ?
X est stationnaire ?

Différence d'ordre 1 :

$$(I - B)X_t = ?$$

La série obtenue est-elle stationnaire ?

Si X est un processus « tendance polynomiale d'ordre 2 », comment stationnariser la série ?

Modèles Intégrés (2/3)

Soit X, processus « saisonnier stable » :

$$X_t = S_t + \varepsilon_t \quad \text{avec} \quad \forall t,\, S_t = S_{t+s}$$

X stationnaire ?

Modèles Intégrés (2/3)

Soit X, processus « saisonnier stable » :

$$X_t = S_t + \varepsilon_t \quad \text{avec} \quad \forall t,\ S_t = S_{t+s}$$

X stationnaire ?

Différence d'ordre 1, avec retard d'ordre s :

$$(I - B^s)X_t = ?$$

La série obtenue est-elle stationnaire ?

Si X comportait en plus une tendance linéaire, que donnerait cette différenciation ?

Modèles Intégrés (3/3)

Une différenciation « simple » d'ordre d supprime les tendances polynomiales d'ordre d :

$$(I - B)^d X_t$$

Une différenciation « saisonnière » supprime aussi les tendances linéaires :

$$(I - B^s) X_t$$

Une différenciation « saisonnière » d'ordre D plus grand que 1 est rare (dans JD+, $D \leq 1$) :

$$(I - B^s)^D X_t$$

Modèles ARIMA

$ARIMA(p,d,q)$ modélise les séries non stationnaires avec tendance

$$\Phi(B)(I-B)^d X_t = \Theta(B)\varepsilon_t$$

$ARIMA(p,d,q)(P,D,Q)$ modélise les séries avec tendance et saisonnalité

$$\Phi(B)\Phi_s(B)(I-B)^d(I-B^s)^D X_t = \Theta(B)\Theta_s(B)\varepsilon_t$$

Factorisation des polynômes en B de la partie *régulière* et de la partie *saisonnière*

Modèles ARIMA et saisonnalité (1/3)

Considérons la partie saisonnière d'un ARIMA :

1 - Une série avec modèle $(p, d, q)(0, 0, 0)$ est-elle saisonnière ?

2 - que dire de $(p, d, q)(0, 0, Q)$?

3 - $(p, d, q)(0, 1, 0)$?

4 - $(p, d, q)(1, 0, 0)$?

Modèles ARIMA et saisonnalité (2/3)

Réponses :

1 - Non, aucune autocorrélation d'ordre s

2 - Non, un MA reflète des fluctuations non persistantes, la saisonnalité persiste dans le temps

3 - Oui, une saisonnalité stable

4 - Ne sait pas, dépend de la valeur du coefficient ϕ_s

- ϕ_s petit en valeur absolue : pas de saisonnalité, phénomène non persistant qui se dissipe vite
- ϕ_s négatif : pas de saisonnalité, phénomène à répétitions bi-annuelles
- ϕ_s grand (proche de 1) et positif : série saisonnière, autocorrélations d'ordre s qui décroît lentement

Modèles ARIMA et saisonnalité (3/3)

Deux cas fréquents :

- $(p, d, q)(0, 1, 1)$ saisonnalité stable en moyenne, avec des fluctuations ponctuelles du niveau de θ_s (plus c'est grand, plus ça fluctue)
- $(p, d, q)(1, 0, 1)$ saisonnalité évolutive avec dérive + fluctuations ponctuelles de niveau θ_s

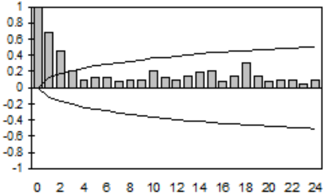
Sommaire

Méthode de Box-Jenkins

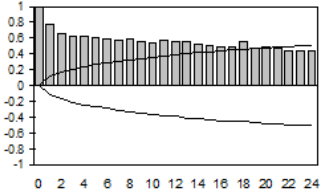
1. Stationnariser le processus : d, D
2. Identifier les ordres ARMA : p, P, q, Q ➔ structure d'autocorrélation de la série
3. Estimer les coefficients ARMA ➔ degré de variabilité de la structure d'autocorrélation
4. Valider le modèle ➔ résidus = bruit blanc ?
5. Choix du modèle (si plusieurs modèles valides) ➔ critères d'information
6. Prévision

Stationnarité et ACF

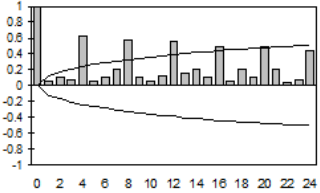
Série stationnaire



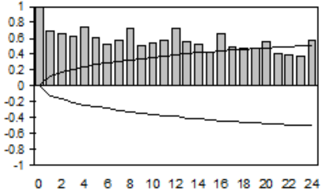
Série non stationnaire : $(I-B)$ **pr station.**



Série non stationnaire : $(I-B^s)$ **pr station.**



Série non stationnaire : $(I-B)(I-B^s)$ **pr sta.**



Choix du modèle

Critères d'information (à **minimiser**) pour comparer les modèles :

- L'AIC (critère de Akaiké) :

$$AIC(p,q) = -2\ln(L) + 2*(p+q)$$

- L'AICC (corrigé pour les courtes périodes) :

$$AICC(p,q) = -2\ln(L) + 2(p+q)\left(1 - \frac{n+p+1}{N_{obs}}\right)^{-1}$$

- Le BIC (critère de Schwarz) :

$$BIC(p,q) = -2\ln(L) + (p+q)\ln(N_{obs})$$

Ne pas comparer des modèles d'ordre de différenciation différents

Sommaire

Principe de TRAMO

TRAMO = Time series Regression with ARIMA noise, Missing values and Outliers

Mêmes objectifs du pré-ajustement de X13-ARIMA (convergence des algorithmes dans JDemetra+ 3.0) :

- corriger la série de points atypiques, des effets de calendrier et imputation des valeurs manquantes
- prolonger la série
- fournir à SEATS le modèle ARIMA à la base de la décomposition

Principe de SEATS (1/3)

SEATS = Signal Extraction in ARIMA Time Series

SEATS utilise le modèle ARIMA de la série linéarisée TRAMO :

$$\underbrace{\Phi(B)\Phi_s(B)(I-B)^d(I-B^s)^D}_{\Phi(B)} X_t = \underbrace{\Theta(B)\Theta_s(B)}_{\Theta(B)} \varepsilon_t$$

Hypothèses :

1. La série linéarisée peut être modélisée par un modèle ARIMA
2. Les différentes composantes sont décorrélées et chaque composante peut être modélisée par un modèle ARIMA
3. Les polynomes AR des composantes n'ont pas de racine commune

Principe de SEATS (2/3)

On factorise le polynôme AR $\Phi(B)$:

$$\Phi(B) = \phi_T(B)\phi_S(B)\phi_C(B)$$

- $\phi_T(B)$ racines correspondant à la tendance
- $\phi_S(B)$ racines correspondant à la saisonnalité
- $\phi_C(B)$ racines correspondant au cycle

Principe de SEATS (3/3)

X_t est exprimé sous la forme :

$$X_t = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t = \underbrace{\frac{\theta_T(B)}{\phi_T(B)}\varepsilon_{T,t}}_{\text{Tendance}} + \underbrace{\frac{\theta_S(B)}{\phi_S(B)}\varepsilon_{S,t}}_{\text{Saisonnalité}} + \underbrace{\frac{\theta_C(B)}{\phi_C(B)}\varepsilon_{C,t}}_{\text{Cycle}} + \underbrace{\nu_t}_{\text{Irrégulier (bruit blanc)}}$$

Un modèle ARIMA est associé à chaque composante.

Infinité de solutions : on retient celle qui minimise la variance de l'irrégulier

➔ Estimation par filtre de Wiener-Kolmogorov

➔ En France c'est X-13ARIMA qui est principalement utilisé (il n'y a pas de "meilleure" méthode)

Sommaire

Les essentiels

Les séries économiques ne sont pas stationnaires, ni leur niveau, ni leurs fluctuations ne sont constants dans le temps

Intégrer un processus permet de le stationnariser

Un MA capte les fluctuations non persistantes autour d'un niveau constant - processus stationnaire

Un AR met en évidence l'influence des réalisations passées sur la réalisation courante

Un ARIMA reflète la structure des autocorrélations de la série, ainsi que le degré de sa variabilité dans le temps

L'examen des résidus permet de valider les modèles, le choix "optimal" se fait grâce aux critères d'information

Exercices

Exercices : écrire les modèles Reg-ARIMA de vos séries à partir des éléments donnés par JDemetra+