

Chapitre 3

Méthodologie de Box-Jenkins

L'approche de Box-Jenkins (1976) consiste en une méthodologie rigoureuse d'étude systématique des séries chronologiques à partir de leur caractéristique. L'objectif est de déterminer le modèle le plus adapté à représenter le phénomène étudié. Cette méthodologie suggère une procédure à trois étapes :

1. Identification du modèle.
2. Estimation des paramètres du modèle.
3. Validation du modèle.

3.1 Identification du modèle

L'étape d'identification d'un processus $ARMA$ (choix entre AR , MA et $ARMA$ et choix de p et q) de BJ est basé sur la comparaison des caractéristiques théoriques à leur équivalent empiriques (c.à.d calculer sur la série observée), les caractéristiques utilisées sont les AC simples et partielles.

-Si le corrélogramme simple n'a que ses q premiers termes différents de zéros et que les termes du corrélogramme partiel diminuent lentement, nous pouvons pronostiquer un $MA(q)$.

-Si le corrélogramme partiel n'a que ses p premiers termes différents de zéros et que les termes du corrélogramme simple diminuent lentement, cela caractérise un $AR(p)$.

-Si les fonctions d'AC simple et partiel ne paraissent pas tronquées, il s'agit d'un processus $ARMA$.

3.1.1 Estimation de d

En pratique si ρ_h est proche de 1 pour un grand nombre de retard c.à.d que la décroissance est lente donc on a une racine unité et le processus n'est pas stationnaire, il faut différencier pour avoir une série stationnaire. d représente le nombre de fois nécessaire pour stationnariser la série. Une autre méthode plus objective est d'utiliser le test de Dickey Fuller.

3.1.2 Estimation de p et q

$$\text{Si } X_t \sim ARIMA(p, d, q) \implies (1 - L)^d X_t \sim ARMA(p, q)$$

*On rappelle l'estimateur de la FAC

$$\hat{\rho}_h = \frac{\sum_{t=h+1}^T (X_t - \hat{\mu})(X_{t-h} - \hat{\mu})}{\sum_{t=1}^T (X_t - \hat{\mu})^2}, \quad h = 1, \dots, T-1.$$

En pratique, on calcule jusqu'à $H = T/4$. Selon Bartlett la variance du coefficient d'AC du processus où $\rho_h = 0$ pour $h > q$ est

$$V(\hat{\rho}_h) = \frac{1}{T} \left(1 + 2 \sum_{j=1}^q \hat{\rho}_j^2 \right)$$

d'où l'intervalle de confiance à 95% des AC

$$\left[\hat{\rho}_h \pm 1.96 \sqrt{\frac{1}{T} \left(1 + 2 \sum_{j=1}^q \hat{\rho}_j^2 \right)} \right]$$

alors $X_t \sim MA(q)$. (En pratique on prend $\pm 1.96 \frac{1}{\sqrt{T}}$).

*La variance des ACP où $\varphi_{hh} = 0$ pour $h > p$ est $\frac{1}{T}$, d'où l'IC à 95% des ACP est

$$\left[\hat{\varphi}_{hh} \pm 1.96 \frac{1}{\sqrt{T}} \right]$$

3.2 Estimation des paramètres du modèle

3.2.1 La méthode des moments

Pour le modèle donné, on sait que les autocorrélations ρ_h et τ_h , des processus AR et MA respectivement, dépendent des paramètres $\xi = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)$ selon des équations théoriques connues

$$(\rho_1, \dots, \rho_p, \tau_1, \dots, \tau_q) = F(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)$$

On sait facilement estimer les autocorrélations ρ_h et τ_h , il suffit donc d'inverser les équations pour estimer les paramètres

$$(\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = F^{-1}(\hat{\rho}_1, \dots, \hat{\rho}_p, \hat{\tau}_1, \dots, \hat{\tau}_q)$$

Exemple

Lors de la modélisation d'une série temporelle on a obtenu : $\hat{\rho}_1 = -0.6, \hat{\rho}_2 = 0.7$ et $\hat{\rho}_3 = -0.5$. On souhaite modéliser la série à l'aide d'un $AR(3)$, proposez des estimateurs des coefficients.

3.2.2 La méthode des moindres carrés (MC)

Modèle AR(p) :

Soit $X_t = \delta + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$ où $\varepsilon_t \sim BB(0, \sigma^2)$.

En donnant N observations x_1, \dots, x_N , les paramètres $\delta, \varphi_1, \dots, \varphi_p$ peuvent être estimés par la méthode des moindres carrés en minimisant

$$S = \sum_{t=p+1}^N (x_t - \delta + \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p})^2$$

par rapport à $\delta, \varphi_1, \dots, \varphi_p$.

Si $\varepsilon_t \sim N \implies$ les estimateurs des MC sont aussi les estimateurs du MV conditionnel sur les p premières valeurs.

Exemple

Soit la série chronologique x_1, \dots, x_N suivant le modèle $AR(1)$ sans constante. Estimer le paramètre φ_1 par les MC.

$$x_t = \varphi_1 x_{t-1} + \varepsilon_t, \quad t = 2, n$$

$$S(\varphi_1) = \sum_{t=2}^n \varepsilon_t^2 = \sum_{t=2}^n (x_t - \varphi_1 x_{t-1})^2$$

on cherche le min de $S(\varphi_1)$

$$\left\{ \begin{array}{l} \frac{\partial S(\varphi_1)}{\partial \varphi_1} = -2 \sum_{t=2}^n x_{t-1} (x_t - \varphi_1 x_{t-1}) = 0 \implies \hat{\varphi}_1 = \frac{\sum_{t=2}^n x_{t-1} x_t}{\sum_{t=2}^n x_{t-1}^2} \\ \frac{\partial^2 S(\varphi_1)}{\partial \varphi_1^2} = 2 \sum_{t=2}^n x_{t-1}^2 > 0. \end{array} \right.$$

Modèle MA :

Le problème d'estimation est plus difficile pour le modèle MA , on ne peut pas trouver des estimateurs explicitement on obtient une forme itérative numérique.

Exemple

Soit le modèle $MA(1)$: $X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$, la somme des résidus aux carrés n'est pas fonction quadratique, on donne l'approche suggéré par BJ :

On commence par des valeurs initiales pour μ et θ_1 , par exemple $\mu = \overline{X}$ et θ_1 donné par l'estimateur des moments, ensuite on calcule la SC des R récursivement à partir de la formule précédente :

$$\varepsilon_t = X_t - \mu + \theta_1 \varepsilon_{t-1} \text{ avec } \varepsilon_0 = 0$$

on a

$$\begin{aligned} \varepsilon_1 &= X_1 - \mu \\ \varepsilon_2 &= X_2 - \mu + \theta_1 \varepsilon_1 \\ &\vdots \\ \varepsilon_N &= X_N - \mu + \theta_1 \varepsilon_{N-1} \end{aligned}$$

alors on calcule $\sum_{t=1}^n \varepsilon_t^2$.

Cette procédure sera répétée pour d'autres valeurs de μ et θ_1 , et $\sum_{t=1}^n \varepsilon_t^2$ sera calculée pour une grille de points dans le plan (μ, θ_1) , ensuite on choisit les *EMC* de μ et θ_1 qui minimise $\sum_{t=1}^n \varepsilon_t^2$.

Remarque

Ces estimateurs sont aussi les *EMVC* avec $\varepsilon_0 = 0$ et $\varepsilon_t \sim N$. Il y a d'autres procédures d'optimisation.

Exercice

Donner la méthode d'estimation des paramètres du modèle *ARMA*(1,1) par la méthode *MCO*.

3.2.3 La méthode du maximum de vraisemblance

Soit le modèle *ARMA* suivant

$$X_t = \delta + \sum_{j=1}^p \varphi_j X_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

avec $\varepsilon_t \sim BB(0, \sigma^2)$. Soit $\vartheta = (\delta, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma^2)$ le vecteur des paramètres.

Soit la série chronologique de taille $n : x_1, \dots, x_n$, l'approche du MV consiste à calculer la densité de probabilité

$$f_{X_N, X_{n-1}, \dots, X_1}(x_n, x_{n-1}, \dots, x_1; \vartheta) \quad (*)$$

L'estimateur du MV de ϑ est la valeur qui maximise (*). On doit spécifier la distribution de ε_t , on suppose que $\varepsilon_t \sim iidN(0, \sigma^2)$.

Trouver l'EMV \implies 2 étapes $\left\{ \begin{array}{l} 1- \text{calculer la fct de V } (*) \\ 2- \text{Trouver les valeurs de } \vartheta \text{ qui maximise cette fct.} \end{array} \right.$

Exemple

Fct de vraisemblance pour un AR(1) gaussien

Processus AR(1) : $X_t = \delta + \varphi_1 X_{t-1} + \varepsilon_t$ tel que $\varepsilon_t \sim iidN(0, \sigma^2)$, le vecteur des paramètres à estimer est $(\delta, \varphi_1, \sigma^2)$, on sait que

$$E(X_1) = \mu = \frac{\delta}{1 - \varphi_1} \quad \text{et} \quad V(X_1) = \frac{\sigma^2}{1 - \varphi_1^2}.$$

$$* \quad \varepsilon_t \sim G \implies X_1 \sim G \implies$$

$$f_{X_1}(x_1; \vartheta) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{\sigma^2}{1 - \varphi_1^2}}} \exp \left(-\frac{\left(x_1 - \frac{\delta}{1 - \varphi_1}\right)^2}{2 \frac{\sigma^2}{1 - \varphi_1^2}} \right)$$

* La distribution de X_2 conditionné par l'observation de X_1

$$X_2 = \delta + \varphi_1 X_1 + \varepsilon_2 \implies X_2/X_1 = x_1 \sim N(\delta + \varphi_1 x_1, \sigma^2).$$

$$f_{X_2/X_1}(x_2/x_1; \vartheta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(x_2 - \delta - \varphi_1 x_1)^2}{2\sigma^2} \right)$$

d'où

$$f_{X_2, X_1}(x_2, x_1; \vartheta) = f_{X_2/X_1}(x_2/x_1; \vartheta) f_{X_1}(x_1; \vartheta)$$

de la même manière

$$f_{X_t/X_{t-1}, \dots, X_1}(x_t/x_{t-1}, \dots, x_1; \vartheta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(x_t - \delta - \varphi_1 x_{t-1})^2}{2\sigma^2} \right)$$

donc la vraisemblance de tout l'échantillon est

$$f_{X_N, X_{n-1}, \dots, X_1}(x_n, x_{n-1}, \dots, x_1; \vartheta) = f_{X_1}(x_1; \vartheta) \prod_{t=2}^n f_{X_t/X_{t-1}}(x_t/x_{t-1}; \vartheta)$$

la fonction de log vraisemblance est, notée $L(\vartheta)$, est

$$L(\vartheta) = \log(f_{X_1}(x_1; \vartheta)) + \sum_{t=2}^n \log(f_{X_t/X_{t-1}}(x_t/x_{t-1}; \vartheta))$$

la valeur ϑ qui maximise $L(\vartheta) \implies$ maximise f , alors

$$\begin{aligned} L(\vartheta) &= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log\left(\frac{\sigma^2}{1-\varphi_1^2}\right) - \frac{\left(x_1 - \frac{\delta}{1-\varphi_1}\right)^2}{2\frac{\sigma^2}{1-\varphi_1^2}} - \frac{n-1}{2} \log(2\pi) \\ &\quad - \frac{n-1}{2} \log(\sigma^2) - \sum_{t=2}^n \frac{(x_t - \delta - \varphi_1 x_{t-1})^2}{2\sigma^2}. \end{aligned}$$

*** L'EMV exacte pour AR(1) gaussien :**

L'EMV $\hat{\vartheta}$ est la valeur pour laquelle $L(\vartheta)$ est maximisé. En principe on dérive par rapport aux paramètres et on annule, on obtient un système d'équation non linéaire, donc pas de solutions analytique et on maximise en utilisant des procédures d'optimisation numérique.

***EMV conditionnel**

On suppose la valeur x_1 déterministe et on maximise

$$f_{X_N, \dots, X_2/X_1}(x_n, \dots, x_2/x_1; \vartheta) = -\frac{n-1}{2} \log(2\pi) - \frac{n-1}{2} \log(\sigma^2) - \sum_{t=2}^n \frac{(x_t - \delta - \varphi_1 x_{t-1})^2}{2\sigma^2}.$$

Maximiser f par rapport à δ et $\varphi_1 \iff$ minimiser $\sum_{t=2}^n (x_t - \delta - \varphi_1 x_{t-1})^2$ c.à.d on obtient les

E des MC. Pour σ^2 on a :

$$\begin{aligned} \frac{\partial f}{\partial \sigma^2} &= -\frac{n-1}{2\sigma^2} + \sum_{t=2}^n \frac{(x_t - \delta - \varphi_1 x_{t-1})^2}{2\sigma^4} = 0 \\ \hat{\sigma}^2 &= \sum_{t=2}^n \frac{(x_t - \delta - \varphi_1 x_{t-1})^2}{n-1}. \end{aligned}$$

3.3 Validation

Lors de la détermination des ordres p et q du processus $ARMA(p, q)$, à l'aide du corrélogramme simple et partiel, on peut être amené à sélectionner plusieurs ordres possibles. Après avoir estimé les différents processus possibles, il reste à les valider. la validation passe par un examen des coefficients estimés (ils doivent être significativement différent de 0) et

par un examen des résidus (les résidus estimés doivent suivre un processus bruit blanc $\widehat{\varepsilon}_t \sim \text{BB}(0, \sigma_\varepsilon^2)$ où $\widehat{\varepsilon}_t$ est l'estimateur de l'erreur ε_t).

3.3.1 Test de Student sur les coefficients

*On vérifie d'abord que les racines des polynômes AR et MA ne sont pas égale à 1. On teste la significativité des coefficients par des tests de student.

On calcule donc la statistique de Student du coefficient $t_{\widehat{\varphi}_j} = \frac{\widehat{\varphi}_j}{\sqrt{\widehat{V}[\widehat{\varphi}_j]}}$ que l'on compare à la valeur critique lue dans la table de la loi de Student. La règle de décision est alors :

Si $|t_{\widehat{\varphi}_j}| < t_{1-\frac{\alpha}{2}}$, on accepte l'hypothèse nulle $H_0 : \varphi_j = 0..$

Si $|t_{\widehat{\varphi}_j}| > t_{1-\frac{\alpha}{2}}$, l'hypothèse nulle est rejetée : $\varphi_j \neq 0..$

Où $t_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $(1 - \frac{\alpha}{2})$ de loi de student à $(T - h)$ degrés de liberté, h étant le nombre des paramètres estimés.

3.3.2 Test sur les résidus

Il s'agit de vérifier que les résidus du modèle estimé, noté $\widehat{\varepsilon}_t$, vérifient les propriétés requises pour que l'estimation soit valide c.à.d : qu'ils suivent un processus bruit blanc non autocorrélé et de même variance et qu'ils suivent une loi normale.

*D'abord il faut regarder le graphique des $\widehat{\varepsilon}_t$ pour voir s'il apparait des points aberrants, une tendance, une rupture, de l'autocorrélation etc, mais ceci n'est qu'indicatif.

Test d'autocorrélation

Analyse des FAC Regarder les AC simples et partielles, elles doivent être significativement égal à 0 si les résidus sont un BB .

$$\widehat{\rho}_h = \frac{\widehat{\gamma}_h}{\widehat{\gamma}_0}, \text{ l'IC dans le cas d'un } BBG : \left[\pm 1.96/\sqrt{T} \right]$$

Pour un BB il faut qu'aucune valeur du corrélogramme ne soit significativement différente de 0.

Test de Ljung-Box(1978) :

$$H_0 : \rho_1 = \rho_2 = \dots \rho_h = 0 \quad \text{VS} \quad H_1 : \exists \text{ un } \rho_i \neq 0$$

La statistique du test est

$$Q_{LB}(h) = T(T+2) \sum_{h=1}^H \frac{\hat{\rho}_h^2}{T-h}$$

Sous l'hypothèse nulle d'absence d'autocorrélation :

$$\hat{\rho}_1^2 = \hat{\rho}_2^2 = \dots = \hat{\rho}_h^2 = 0$$

La statistique Q_{LB} suit une loi de Khi-Deux à $(H - p - q)$ degrés de liberté.

Tests d'hétéroscédasticité

Nous étudierons ici le test ARCH car il est très fréquemment employé en économétrie des séries temporelles financières.

Test ARCH

Ce test consiste à effectuer une régression autorégressive des résidus carrés sur q retard :

$$e_t^2 = \theta_0 + \sum_{j=1}^q \theta_j e_{t-j}^2$$

où e_t désigne le résidu à l'instant t issu de l'estimation des paramètres du processus $ARMA(p, q)$. Pour déterminer le nombre de retards q , on étudie le corrélogramme des résidus au carré.

Les hypothèses du test *ARCH* sont les suivantes :

$$\begin{cases} H_0 : \text{homoscédasticité} : \theta_1 = \dots = \theta_q = 0 \\ H_1 : \text{hétéroscédasticité} : \exists \text{ un } \theta_i \neq 0 \end{cases}$$

La statistique du test est $T \times R^2$ où T correspond au nombre d'observations de la série e_t et R^2 représente le coefficient de détermination associé à la régression

Sous l'hypothèse H_0 la statistique $T \times R^2$ suit la loi du Khi-deux à q degrés de liberté.

La règle de décision est alors :

- Si $T \times R^2 \leq \chi^2_q$ où χ^2_q on accepte H_0 d'homoscédasticité.
- Si $T \times R^2 > \chi^2_q$ où χ^2_q on rejette H_0 et on admet qu'il y a de l'hétéroscédasticité.

Tests de normalité

On utilise le test de Jarque et Bera (1984) basé sur la notion de skewness (coefficient d'asymétrie) et de Kurtosis (aplatissement-épaisseur des queues de distribution). Soit μ_k le moment empirique d'ordre k du processus $\hat{\varepsilon}_t : \mu_k = E [\hat{\varepsilon}_t - E [\hat{\varepsilon}_t]]^k = \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_t - \bar{\varepsilon}_t)^k$

Les coefficients de la Skewness (S_k) et de la Kurtosis (K_u) est alors définie par :

$$(S_k)^{\frac{1}{2}} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} \longrightarrow_{T \rightarrow \infty} \mathcal{N} \left(0, \sqrt{\frac{6}{T}} \right)$$

$$K_u = \frac{\mu_4}{\mu_2^2} \longrightarrow_{T \rightarrow \infty} \mathcal{N} \left(3, \sqrt{\frac{24}{T}} \right)$$

On construit alors les statistiques centrées réduites correspondantes à $(S_k)^{\frac{1}{2}}$ et K_u , qu'on compare aux seuils d'une loi normale centrée réduite. La statistique est calculée comme suit :

$$\frac{(S_k)^{\frac{1}{2}}}{\sqrt{\frac{6}{T}}} \longrightarrow_{T \rightarrow \infty} \mathcal{N} (0, 1)$$

$$\frac{K_u - 3}{\sqrt{\frac{24}{T}}} \longrightarrow_{T \rightarrow \infty} \mathcal{N} (0, 1)$$

Si la statistique centrée réduite de $(S_k)^{\frac{1}{2}}$ est inférieure au seuil 1,96 à 5%, on accepte l'hypothèse de symétrie et l'hypothèse de normalité. Si la statistique centrée réduite de K_u est inférieure au seuil 1,96 à 5%, on accepte l'hypothèse de queue de distributions plates et l'hypothèse de normalité.

Le test de Jarque et Bera regroupe ces deux tests en un seul test. On construit la statistique :

$$S = \frac{T}{6} S_k + \frac{T}{24} (K_u - 3)^2 \xrightarrow{T \rightarrow \infty} \chi_2^2 = 5.99$$

Donc si $S \geq \chi_2^2 (1 - \alpha)$ on rejette l'hypothèse H_0 de normalité des résidus au seuil de $\alpha\%$.

3.3.3 Critères de choix des modèles

Après examen des coefficients et des résidus, certains modèles sont écartés. Pour départager les modèles restants, on fait appel aux critères standards et critères d'information.

Critères standards

L'erreur absolue moyenne (MAE)

$$MAE = \frac{1}{T} \sum_t |e_t|$$

Racine de l'erreur quadratique moyenne (RMSE)

$$RMSE = \sqrt{\frac{1}{T} \sum_t e_t^2}$$

Ecart absolu moyen en pourcentage (MAPE)

$$MAPE = 100 \frac{1}{T} \sum_t \left| \frac{e_t}{X_t} \right|$$

Plus la valeur de ces critères est faible, plus le modèle estimé est proche des observations.

Critère d'information

* AIC (Akaike) :

$$AIC(p, q) = \log \hat{\sigma}^2 + \frac{2(p + q)}{T}$$

*Le critère de Schwarz (1978) :

$$SC(p, q) = \log(\sigma_{\varepsilon_t}^2) + (p + q) \frac{\log T}{T}$$

On choisit le modèle qui minimise les critères standards et les critères d'information. Le modèle sélectionné sera alors utilisé pour la prévision

3.4 Prévision

Après validation du modèle on procède à des prévisions sur un horizon limité car la variance de l'erreur de prévision croît rapidement avec l'horizon.

Si l'on désire avoir la valeur de X_{t+h} , tout en supposant qu'on est à l'instant t , on utilise l'estimateur $\widehat{X}_t(h) = E(X_{t+h}/X_t, X_{t-1}, \dots, X_1)$ avec t l'origine de prévision et h l'horizon de prévision.

dans le cas d'un modèle *ARMA*, on aura la forme :

$$\begin{aligned} \widehat{X}_t(h) &= \varphi_1 E(X_{t+h-1}/X_t, X_{t-1}, \dots, X_1) + \dots + \varphi_p E(X_{t+h-p}/X_t, X_{t-1}, \dots, X_1) \\ &\quad - \theta_1 E(\varepsilon_{t+h-1}/X_t, X_{t-1}, \dots, X_1) - \dots - \theta_q E(\varepsilon_{t+h-q}/X_t, X_{t-1}, \dots, X_1) + E(\varepsilon_{t+h}/X_t) \end{aligned}$$

avec :

$$\begin{cases} E(X_{t-j}/X_t, X_{t-1}, \dots, X_1) = X_{t-j}, & \forall j \geq 0 \\ E(X_{t+j}/X_t, X_{t-1}, \dots, X_1) = \widehat{X}_t(j), & \forall j \geq 1 \\ E(\varepsilon_{t-j}/X_t, X_{t-1}, \dots, X_1) = \varepsilon_{t-j}, & \forall j \geq 0 \\ E(\varepsilon_{t+j}/X_t, X_{t-1}, \dots, X_1) = 0 & \forall j \geq 1 \end{cases}$$

Erreur de prévision

On considère un processus *ARMA*(p, q) telle que :

$$X_t = \delta + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

avec $(\varphi_p, \theta_q) \in \mathbb{R}^{*2}$ et $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$

Appliquons le théorème de Wold au processus $(X_t, t \in \mathbb{Z})$ et considérons la forme $MA(\infty)$ correspondante : (avec $\delta = 0$)

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}; \psi_0 = 1$$

Il s'ensuit que la meilleur prévision que l'on peut faire de X_{t+1} compte tenu de toute l'information disponible jusqu'à la date t notée $\widehat{X}_t(1)$ est donnée par :

$$\begin{aligned}\widehat{X}_t(1) &= E(X_{t+1}/X_t, X_{t-1}, \dots, X_1) \\ &= E(X_{t+1}/\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_1) \\ &= \sum_{j=1}^{\infty} \psi_j \varepsilon_{t+1-j}\end{aligned}$$

L'erreur de prévision ε_t est appelé aussi l'innovation :

$$X_{t+1} - \widehat{X}_t(1) = \varepsilon_{t+1}$$

plus généralement pour une prévision à un horizon k on a :

$$\begin{aligned}\widehat{X}_t(k) &= \sum_{j=k}^{\infty} \psi_j \varepsilon_{t+k-j} \\ X_{t+k} - \widehat{X}_t(k) &= \sum_{j=0}^{k-1} \psi_j \varepsilon_{t+k-j}\end{aligned}$$

Déterminons un intervalle de confiance sur la prévision $\widehat{X}_t(k)$, sous l'hypothèse de normalité des résidus ε_t . On montre alors que :

$$E\{[X_{t+k} - \widehat{X}_t(k)]^2\} = E\left\{\left[\sum_{j=0}^{k-1} \psi_j \varepsilon_{t+k-j}\right]^2\right\} = \sum_{j=0}^{k-1} \psi_j^2 \sigma_\varepsilon^2$$

D'où :

$$\frac{X_{t+k} - \widehat{X}_t(k)}{\sigma_\varepsilon \left[\sum_{j=0}^{k-1} \psi_j^2\right]^{1/2}} \sim N(0, 1)$$

On peut donc construire un intervalle de confiance sous la forme :

$$IC = \left[\widehat{X}_t(k) \pm t_{\frac{\alpha}{2}} \widehat{\sigma} \left[\sum_{j=0}^{k-1} \psi_j^2 \right]^{1/2} \right].$$

Exemples

1) Donner la meilleure prévision du processus $AR(1)$ à l'horizon h et l'erreur prévisionnelle.

Application : $X_t = 5 + 0.5X_{t-1} + \varepsilon_t/\varepsilon_t \sim N(0, 1)$, IC de X_{t+1} et X_{t+2} avec la valeur 10.738 comme dernière observation.

2) Même question pour $MA(1)$. Application : les dernières observations : 5.654 ; 4.686 ; 5.965.

Bibliographie

- [1] Brockwell, J. P. and Davis, R. A. (2002). Introduction to Time Series and Forecasting. 2nd ed. Springer texts in statistics. Springer-Verlag New York, Inc.
- [2] Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press. Princeton, New Jersey.
- [3] Kirchgässner, G. and Wolters, J. (2007). Introduction to modern time series analysis. ISBN 978-3-540-73290-7 Springer Berlin Heidelberg New York.