

U.F.R. Economie Appliquée

Maîtrise d'Economie Appliquée

Cours de Tronc Commun

Econométrie Appliquée

Séries Temporelles

Christophe HURLIN

Chapitre 4

Estimation, Tests de Validation et Prévisions des Processus ARMA

La procédure de modélisation de Box et Jenkins (1976) comporte les étapes suivantes :

- Stationnarisation et Dessaisonalisation
- Identification
- Estimation
- Validation et Test
- Prévisions

A la suite des chapitres précédents, reste à étudier les 3 derniers points.

1. Estimation

L'estimation des paramètres d'un modèle $ARMA(p, q)$ lorsque les ordres p et q sont supposés connus peut se réaliser par différentes méthodes dans le domaine temporel :

- Moindres Carrés Ordinaires (modèle sans composante MA , $q = 0$). Dans ce cas, on retrouve les équations de Yule Walker. En remplaçant les autocorrélations théoriques par leurs estimateurs, on peut retrouver les estimateurs des MCO des paramètres du modèle par la résolution des équations de Yule Walker.
- Maximum de Vraisemblance approché (Box and Jenkins 1970)
- Maximum de Vraisemblance exacte (Newbold 1974, Harvey et Philips 1979, Harvey 1981)

Nous allons présenter ici brièvement la démarche de l'estimation par le maximum de vraisemblance. Cette maximisation est réalisée à l'aide d'algorithmes d'optimisation non linéaire (Newton-Rahpson, méthode du simplex) que nous n'exposerons pas dans le cadre de ce chapitre. Nous nous contenterons ici de montrer comment s'écrit le programme de maximisation de la vraisemblance permettant d'estimer les paramètres d'un modèle $ARMA(p, q)$.

1.1. Rappels : l'estimateur du maximum de vraisemblance

Definition 1.1. Soit X une variable aléatoire à valeur dans (\mathcal{X}, a) de loi P_θ . On note $f(x, \theta)$ la densité de P_θ et $f(x_1, \dots, x_n, \theta)$ la densité empirique correspondante. On appelle vraisemblance du paramètre θ l'application $\Theta \rightarrow \mathbb{R}^+$ définie par :

$$\forall \theta \in \Theta \rightarrow L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (1.1)$$

Definition 1.2. Soit $L(x, \theta)$ la vraisemblance au point θ , $\theta \in \mathcal{Z}$. On appelle estimateur du maximum de vraisemblance pour θ la statistique:

$$\begin{aligned} \hat{\theta} &: \mathcal{X}^T \rightarrow \mathcal{Z} \\ (x_1, \dots, x_n) &\rightarrow \hat{\theta}(x_1, \dots, x_n) \end{aligned}$$

telle que :

$$\forall \theta \in \Theta \quad L(x, \hat{\theta}) \geq L(x, \theta)$$

Le principe de la vraisemblance revient à déterminer la valeur du paramètre θ , fonction des observations (x_1, \dots, x_n) , qui assure la plus grande probabilité d'apparition de ces observations (x_1, \dots, x_n) .

Corollary 1.3. Lorsque l'on suppose que (i) l'ensemble \mathcal{X} est indépendant de θ et que (ii) la fonction de vraisemblance $L(\cdot)$ est deux fois continûment différentiable par rapport à θ , $\forall \theta \in \Theta$ alors l'estimateur du maximum de vraisemblance $\hat{\theta}$ est solution du système:

$$\left(\frac{\partial L(x, \theta)}{\partial \theta} \right)_{\theta=\hat{\theta}} = 0 \quad (1.2)$$

$$\left(\frac{\partial^2 L(x, \theta)}{\partial \theta^2} \right)_{\theta=\hat{\theta}} < 0 \quad (1.3)$$

Theorem 1.4. S'il existe un estimateur efficace du paramètre θ (au sens de la borne de Cramer Rao), alors cet estimateur est identique à celui du maximum de vraisemblance $\hat{\theta}$.

1.2. Application aux modèles ARMA

On considère un processus stationnaire $\{x_t\}$ satisfaisant une représentation $ARMA(p, q)$ telle que :

$$\Phi(L)x_t = c + \Theta(L)\varepsilon_t \quad (1.4)$$

avec $c \in \mathbb{R}$, $\Theta(L) = \sum_{j=0}^q \theta_j L^j$, $\Phi(L) = \sum_{j=0}^p \phi_j L^j$ où $\forall j < q \theta_j \in \mathbb{R}^2, \forall j < p \phi_j \in \mathbb{R}^2$, $\theta_0 = \phi_0 = 1$ et $(\phi_p, \theta_q) \in \mathbb{R}^{2*}$.

En plus de la définition standard d'un processus *ARMA*, on fait l'hypothèse de la normalité des résidus afin de spécifier une forme fonctionnelle à la vraisemblance du modèle.

Hypothèse H1 On suppose que la population des résidus $\{\varepsilon_t\}$ peut être décrite par un processus bruit blanc gaussien $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Ecrivons alors la vraisemblance associée au vecteur de réalisation (x_1, x_2, \dots, x_T) .

$$(2\pi\sigma_\varepsilon^2)^{-T/2} \det [\Omega(\theta_i, \phi_i)]^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} x' [\Omega(\theta_i, \phi_i)]^{-1} x \right\}$$

2. Tests de Validation

2.1. Tests de redondance

Le but est de vérifier si les composantes *AR* et *MA* de l'*ARMA* n'ont pas de racines communes. Si tel est le cas, on peut alors se ramener à une représentation minimale excluant ces racines. Cette représentation sera préférable selon le principe de parcimonie.

Exemple : On considère un processus stationnaire $\{x_t, t \in \mathbb{Z}\}$ satisfaisant une représentation *ARMA* (p, q) telle que :

$$\Phi(L) x_t = \Theta(L) \varepsilon_t$$

Soient $\lambda_i \in \mathbb{C}$, $i \in [1, \tilde{p}]$, $\tilde{p} \leq p$ les racines de $\Phi(L) = 0$ et soient $\mu_i \in \mathbb{C}$, $i \in [1, \tilde{q}]$, $\tilde{q} \leq q$ les racines de $\Theta(L) = 0$. Supposons qu'il existe une racine commune à ces deux polynômes.

$$\exists j \in [1, \min(\tilde{p}, \tilde{q})] \text{ tel que } \lambda_j = \mu_j$$

Alors on peut réexprimer les deux polynômes sous la forme :

$$\begin{aligned} \Phi(L) &= \prod_{i=1}^{\tilde{p}} \left(1 - \frac{L}{\lambda_i}\right) = \left(1 - \frac{L}{\lambda_j}\right) \prod_{i=1, i \neq j}^{\tilde{p}} \left(1 - \frac{L}{\lambda_i}\right) \\ \Theta(L) &= \prod_{i=1}^{\tilde{q}} \left(1 - \frac{L}{\mu_i}\right) = \left(1 - \frac{L}{\mu_j}\right) \prod_{i=1, i \neq j}^{\tilde{q}} \left(1 - \frac{L}{\mu_i}\right) \end{aligned}$$

Dès lors, en divisant les deux polynômes par $(1 - L\lambda_j^{-1}) = (1 - L\mu_j^{-1})$, le processus $\{x_t, t \in \mathbb{Z}\}$ peut se réécrire sous la forme :

$$\tilde{\Phi}(L) x_t = \tilde{\Theta}(L) \varepsilon_t \Leftrightarrow \prod_{i=1, i \neq j}^{\tilde{p}} \left(1 - \frac{L}{\lambda_i}\right) x_t = \prod_{i=1, i \neq j}^{\tilde{q}} \left(1 - \frac{L}{\mu_i}\right) \varepsilon_t \quad (2.1)$$

La représentation (2.1) correspond à la représentation minimale du processus $\{x_t, t \in \mathbb{Z}\}$. Bien entendu, dans cette représentation les degrés de la représentation *ARMA* seront strictement inférieurs à ceux de la représentation initiale, d'où un gain de degré de liberté au moment de la phase d'estimation.

2.2. Tests de significativité des coefficients

Quelle que soit la méthode d'estimation employée, il est possible de calculer la matrice de variance covariance des $p + q + 1$ estimateurs des paramètres d'un modèle *ARMA* (p, q) , notés $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$. Supposons que σ_ε^2 est connu. En particulier pour l'estimateur du maximum de vraisemblance on a le résultat suivant :

Theorem 2.1. *L'estimateur du maximum de vraisemblance du vecteur des paramètres du modèle *ARMA* (p, q) , noté $\widehat{\Delta} = (\widehat{\phi}_1, \dots, \widehat{\phi}_p, \widehat{\theta}_1, \dots, \widehat{\theta}_q)$, est asymptotiquement distribué suivant une loi normale de moyenne Δ et de variance covariance*

$$E(\widehat{\Delta}\widehat{\Delta}') = \mathcal{F}^{-1} \quad (2.2)$$

où \mathcal{F} désigne la matrice d'information de Fischer

$$\mathcal{F} = -E \frac{\partial^2 L(x, \Delta)}{\partial \delta_i \partial \delta_j} \quad (\delta_i, \delta_j) \in \Delta^2 \quad (2.3)$$

Partant de la matrice de variance covariance des estimateurs du maximum de vraisemblance, il est possible de reconstruire les statistiques de Student associées aux différents paramètres du modèle *ARMA*. Sur le plan appliqué, lorsque le processus $\{x_t, t \in \mathbb{Z}\}$ est stationnaire, on peut montrer que la distribution asymptotique de ces statistiques de Student sont asymptotiquement distribuées selon une loi normale. Donc, on peut appliquer à ces estimateurs les méthodes d'inférence traditionnelles.

Remark 1. *Si l'on montre que l'un ou plusieurs paramètres du modèle ne sont pas significativement différents de 0, on estime à nouveau le modèle en excluant les variables correspondantes (erreur de spécification).*

2.3. Coefficient de détermination

Tout comme dans le cas des modèles linéaires standard, le coefficient de détermination donne une information sur la part de la variance de la variable endogène (ici x_t) qui peut être expliquée par le modèle estimé.

Rappels On considère un processus stationnaire $\{x_t, t \in \mathbb{Z}\}$ satisfaisant une représentation $ARMA(p, q)$. On note $\hat{\varepsilon}_t$ le résidu d'estimation du modèle. Les coefficients de détermination R^2 et \bar{R}^2 sont alors définis par :

$$R^2 = 1 - \frac{\sum_{t=1}^T \hat{\varepsilon}_t}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

$$\bar{R}^2 = 1 - \frac{T-1}{T-p-q} \frac{\sum_{t=1}^T \hat{\varepsilon}_t}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

On utilise de préférence le \bar{R}^2 puisqu'il permet de prendre en compte le nombre de variables explicatives, c'est à dire les p termes retardés de l' AR et les q retards de la composante MA . Bien entendu ces coefficients sont proches de 1 lorsque l'ajustement du modèle aux données est parfaite, c'est à dire si $\sum_{t=1}^T \hat{\varepsilon}_t$ tend vers 0.

2.4. Test de Bruit Blanc

Lorsque le processus est bien estimé, les résidus entre les valeurs observées et les valeurs estimées par le modèle doivent se comporter comme un bruit blanc. On notera par la suite $\hat{\varepsilon}_t$ le résidu d'estimation du modèle.

2.4.1. Test de nullité de la moyenne des résidus

Soit T le nombre de données disponibles (après avoir enlevé les retards correspondant aux termes AR et MA). Si le processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ est *i.i.d.* $(0, \sigma_\varepsilon^2)$, on doit avoir:

$$\bar{\varepsilon}_t = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \xrightarrow{T \rightarrow \infty} 0$$

Par application du théorème central limite, on montre que :

$$\frac{\bar{\varepsilon}_t}{\hat{\sigma}_{\varepsilon_t}} \sqrt{T} \xrightarrow{T \rightarrow \infty} \mathcal{N}(0, 1)$$

Dès lors, on peut tester la nullité de la moyenne des résidus en construisant l'intervalle de confiance sur $\bar{\varepsilon}_t$ au seuil standard de 95%.

$$P \left\{ \bar{\varepsilon}_t \in \left[\frac{-1.96 \hat{\sigma}_{\varepsilon_t}}{\sqrt{T}}, \frac{1.96 \hat{\sigma}_{\varepsilon_t}}{\sqrt{T}} \right] \right\} = 0.95$$

2.4.2. Test d'autocorrélation des résidus

Si les résidus $\{\varepsilon_t, t \in \mathbb{Z}\}$ obéissent à un bruit blanc, il ne doit pas exister d'autocorrélation dans la série. On peut alors utiliser les différents tests suivants :

1. Test de Durbin Watson : test de l'autocorrélation d'ordre 1.
2. Etude de la FAC et de la FAP : on doit vérifier qu'il n'existe aucune autocorrélation ou autocorrélation partielle significativement non nulle pour le processus étudié. Cette étude est prolongée par les tests du "porte-manteau"
3. Tests du "porte-manteau" ou tests d'adéquation globale du modèle. Ces tests reposent sur l'idée que la FAC d'un bruit blanc ne doit pas révéler d'autocorrélations non nulles. En pratique, on utilise deux tests :

Test de Box et Pierce :

On note r_k l'autocorrélation d'ordre k du processus $\{\varepsilon_t, t \in \mathbb{Z}\}$. Pour un ordre K , le test de Box et Pierce est le test de l'hypothèse $H_0 : r_1 = \dots = r_K = 0$ contre $H_1 : \exists j \in [1, K]$, tel que $r_j \neq 0$. Pour un processus $ARMA(p, q)$, la statistique de ce test est :

$$Q_{BP} = T \sum_{k=1}^K r_k^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \chi^2(K - p - q)$$

L'hypothèse H_0 est rejetée au seuil de 5% si Q_{BP} est supérieur au quantile 0.95 de la loi du χ^2 correspondant.

Test de Ljung-Box¹ :

Ces statistiques, définies pour un ordre K , correspondent à l'hypothèse nulle $H_0 : r_k = 0 \forall k \leq K$ et sont construites de la façon suivante :

$$Q_K = T(T+2) \sum_{k=1}^K \frac{r_k^2}{T-k} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \chi^2(K - p - q)$$

1. Test de Von Neumann's : on range par ordre croissant les valeurs des résidus. Soit R_i la nouvelle chronique obtenue, le coefficient du Von Neumann's ratio test est

$$RVN = \frac{\sum_{t=1}^{T-1} (R_t - R_{t+1})^2}{\sum_{t=1}^{T-1} (R_t - \bar{R})^2}$$

où \bar{R} est la moyenne des R_i . L'hypothèse H_0 d'indépendance des valeurs des résidus, est rejetée si $RVN > \tau$ où τ est la valeur critique de Bartels (à 95%, $\tau = 1.67$).

¹Sous Eviews noté Q-stats

2. Test du "CUSUM" : Ce test permet d'étudier la stabilité du modèle estimé au cours du temps. Il existe deux versions de ce test : le CUSUM fondé sur la somme cumulée des résidus récurrents et le CUSUM SQ fondé sur la somme cumulée du carré des résidus récurrents. On note $\tilde{\varepsilon}_t$ le résidu normalisé par son écart type, tel que $\tilde{\varepsilon}_t = \hat{\varepsilon}_t / \hat{\sigma}_\varepsilon$. On note k le nombre de paramètres à estimer du modèle. Soit la statistique S_t du CUSUM et la statistique S'_t du CUSUM SQ, on a :

$$S_t = (T - k) \frac{\sum_{j=k+1}^t \tilde{\varepsilon}_j}{\sum_{j=k+1}^t \tilde{\varepsilon}_j^2} \quad t = k + 1, \dots, T$$

$$S'_t = \frac{\sum_{j=k+1}^t \tilde{\varepsilon}_j^2}{\sum_{j=k+1}^T \tilde{\varepsilon}_j^2} \quad t = k + 1, \dots, T$$

Si les coefficients sont stables au cours du temps, alors les résidus récurrents S_t doivent rester dans l'intervalle défini par

$$\left[\pm \frac{\alpha (2t + T - 3k)}{\sqrt{T - k}} \right]$$

où $\alpha = 1.143, 0.948, 0.850$ pour des seuils respectivement égaux à 1%, 5% et 10%. De la même façon, les résidus S'_t doivent être compris dans l'intervalle

$$\left[\pm C \frac{(t - T)}{T - k} \right]$$

où C est la statistique de Kolmogorov-Smirnov.

2.4.3. Test d'homoscédasticité

Un bruit blanc est par définition homoscédastique. Tous les tests d'hétéroscédasticité peuvent ici être employés. Test de Chow (comparaison des variances des résidus sur des sous-périodes de la chronique),

2.5. Tests de normalité

Pour vérifier si le processus des résidus $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc gaussien, plusieurs tests peuvent être utilisés, mais le test le plus courant est celui de Jarque et Bera. Ce dernier est fondé sur la notion de skewness (moment d'ordre 3 et asymétrie) et de Kurtosis (moment

d'ordre 4 et queue de distribution). Soit μ_k le moment empirique d'ordre k du processus $\{\varepsilon_t, t \in \mathbb{Z}\}$:

$$\mu_k = \frac{1}{T} \sum_{i=1}^t (\hat{\varepsilon}_t - \bar{\varepsilon}_t)^k$$

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

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

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

On construit alors les statistiques centrées réduites correspondantes à $S_k^{1/2}$ et K_u que l'on compare aux seuils d'une loi normale centrée réduite.

$$\frac{(S_k)^{1/2}}{\sqrt{\frac{6}{T}}} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

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

Si la statistique centrée réduite de $S_k^{1/2}$ 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 :

$$s = \frac{T}{6} S_k + \frac{T}{24} (K_u - 3)^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \mathcal{X}^2(2)$$

Donc si $s \geq \mathcal{X}_{1-\alpha}^2(2)$ on rejette l'hypothèse H_0 de normalité des résidus au seuil de $\alpha\%$.

2.6. Critères de comparaison des modèles

Au delà des critères standard (MSE, MAE, RMSE, FPE etc.), on étudiera les critères propres aux modèles autorégressifs.

1. Critère de Akaike ou AIC : Le meilleur des modèles $ARMA(p, q)$ est le modèle qui minimise la statistique :

$$AIC(p, q) = T \log(\sigma_{\hat{\varepsilon}_t}^2) + 2(p + q) \quad (2.4)$$

2. Le critère d'information bayésien (ou BIC) : ce critère présente l'avantage de plus pénaliser les modèles où les paramètres sont en surnombre comparativement à l'AIC

$$BIC(p, q) = T \log(\sigma_{\hat{\varepsilon}_t}^2) - (n - p - q) \log \left[1 - \frac{(p + q)}{T} \right] + (p + q) \log(T) + \log \left[(p + q)^{-1} \left(\frac{\sigma_{x_t}^2}{\sigma_{\hat{\varepsilon}_t}^2} - 1 \right) \right] \quad (2.5)$$

3. Le critère de Schwarz (1978) :

$$SC(p, q) = T \log(\sigma_{\hat{\varepsilon}_t}^2) + (p + q) \log(T)$$

4. Le critère de Hannan-Quin (1979) :

$$HQ(p, q) = \log(\sigma_{\hat{\varepsilon}_t}^2) + (p + q) c \log \left[\frac{\log(T)}{T} \right]$$

où c est une constante à spécifier.

3. Prévision

3.1. Transformation de la série

Lorsque pour identifier le processus étudié à un processus *ARMA*, on a appliqué différentes transformations (exemple différenciation dans le cas d'une série $I(1)$), il est nécessaire lors de la phase de prévision de prendre en compte la transformation retenue et de "recoler la prévision". Plusieurs cas sont possibles :

- Si le processus contient une tendance déterministe, on extrait cette dernière par régression afin d'obtenir une série stationnaire lors de la phase d'estimation. Ensuite, lors de la phase de prévision, on adjoint aux prévisions réalisées sur la composante *ARMA* stationnaire, la projection de la tendance.
- Si la transformation résulte de l'application d'un filtre linéaire (de type par exemple différences premières), on réalise les prévisions sur la série filtrée stationnaire et l'on reconstruit ensuite par inversion du filtre les prévisions sur la série initiale.

3.2. Prédicteur pour un processus *ARMA*

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

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

avec $(\phi_p, \theta_q) \in \mathbb{R}^{2*}$ et ε_t *i.i.d.* $(0, \sigma_\varepsilon^2)$. Appliquons le théorème de Wold au processus $\{x_t, t \in \mathbb{Z}\}$ et considérons la forme $MA(\infty)$ correspondante :

$$x_t = \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-j} \quad \pi_0 = 1 \quad (3.1)$$

Il s'ensuit que la meilleure prévision que l'on peut faire de x_{t+1} compte tenu de toute l'information disponible jusqu'à la date t , notée $\hat{x}_t(1)$, est donnée par :

$$\begin{aligned} \hat{x}_t(1) &= E(x_{t+1} / x_t, x_{t-1}, x_{t-2}, \dots, x_0) \\ &= E(x_{t+1} / \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_0) \\ &= \sum_{j=1}^{\infty} \pi_j \varepsilon_{t+1-j} \end{aligned} \quad (3.2)$$

Dès lors, l'erreur de prévision est donnée par la réalisation en $t+1$ de l'innovation qui en t n'est pas connue :

$$x_{t+1} - \hat{x}_t(1) = \varepsilon_{t+1} \quad (3.3)$$

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

$$\hat{x}_t(k) = \sum_{j=k}^{\infty} \pi_j \varepsilon_{t+k-j} \quad (3.4)$$

$$x_{t+k} - \hat{x}_t(k) = \sum_{j=0}^{k-1} \pi_j \varepsilon_{t+k-j} \quad (3.5)$$

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

$$\frac{x_{t+k} - \hat{x}_t(k)}{\text{var}[x_{t+k} - \hat{x}_t(k)]^{1/2}} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Or on sait que :

$$E\{[x_{t+k} - \hat{x}_t(k)]^2\} = E\left[\left(\sum_{j=0}^{k-1} \pi_j \varepsilon_{t+k-j}\right)^2\right] = \sum_{j=0}^{k-1} \pi_j^2 \sigma_\varepsilon^2$$

D'où

$$\frac{x_{t+k} - \hat{x}_t(k)}{\sigma_\varepsilon \left[\sum_{j=0}^{k-1} \pi_j^2\right]^{1/2}} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

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

$$IC = \left[\hat{x}_t(k) \pm t^{\alpha/2} \left(\sum_{j=0}^{k-1} \pi_j^2 \right)^{1/2} \hat{\sigma}_\varepsilon \right]$$