



SÉRIES TEMPORELLES LINÉAIRES

Projet : Modélisation ARIMA d'une série temporelle

Paul LEMOINE VANDERMOERE

Mihnea POPA

Table des matières

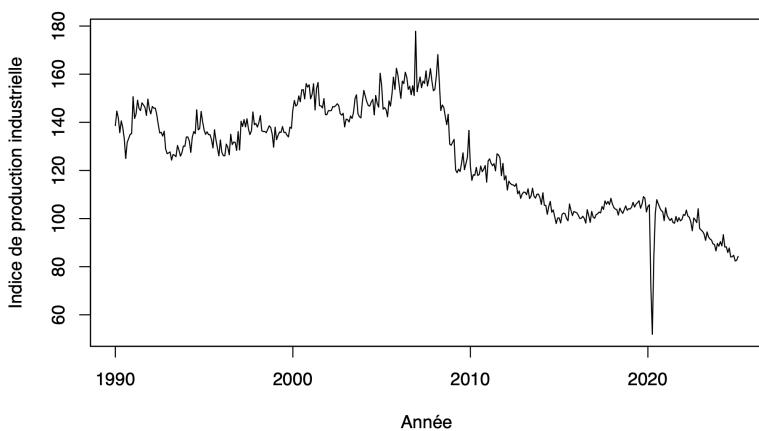
1 Partie I	1
1.1 Question 1	1
1.2 Question 2	1
1.3 Question 3	2
2 Partie II	2
2.1 Question 4	2
2.2 Question 5	4
3 Partie III	4
3.1 Question 6	4
3.2 Question 7	5
3.3 Question 8	5
3.4 Question 9	6
4 Annexes	7

1 Partie I

1.1 Question 1

La série choisie correspond à l'indice de production (base 2021) industrielle dans le secteur du Travail du bois et fabrication d'articles en bois et en liège, à l'exception des meubles, fabrication d'articles en vannerie et sparterie. Elle est caractérisée par une fréquence mensuelle et s'étend de janvier 1990 à février 2025. Il s'agit d'un indice CVS-CJO, ce qui signifie que la série est corrigée des variations saisonnières et des jours ouvrés, donc elle est déjà pré-traitée pour éliminer les effets saisonniers réguliers. Ainsi, la série ne présente pas de saisonnalité au vu de son évolution. Cependant, graphiquement nous remarquons qu'elle n'est pas stationnaire. 1. Il s'agira dans la question suivante de la rendre stationnaire.

FIGURE 1 – Représentation graphique de la série étudiée - janvier 1990 à février 2025



1.2 Question 2

Afin de vérifier la stationnarité de notre série temporelle, nous avons d'abord appliqué une différenciation d'ordre 1. Formellement, soit $(Y_t)_{t \in \mathbb{Z}}$ une série temporelle. Sa première différence est définie par

$$\Delta Y_t = Y_t - Y_{t-1}$$

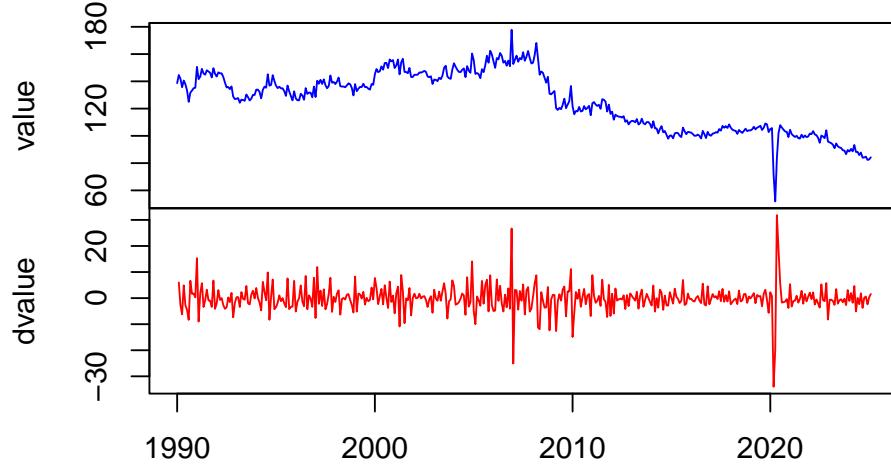
Autrement dit, on transforme la série initiale en une nouvelle série qui mesure variations entre deux observations consécutives. Cela permet de retirer une éventuelle tendance ou non-stationnarité structurelle. À la lecture du graphique 6, la série différenciée *dvalue* ne présente pas de tendance apparente ni de saisonnalité persistante. Cela suggère que la série est stationnaire, ou du moins proche de l'être, c'est-à-dire qu'elle ne possède pas de racine unitaire. Toutefois, une confirmation formelle à l'aide de tests statistiques de racine unitaire est nécessaire.

Pour déterminer le type exact de test à utiliser, nous avons estimé une régression linéaire de *dvalue* sur le temps. Cette régression ne montre ni constante significative ni tendance linéaire, ce qui correspond au cas sans dérive ni tendance. Cela nous oriente vers un test de Phillips-Perron (PP) avec l'hypothèse alternative selon laquelle la série est stationnaire. Le test PP donne une p-value de 0.01, ce qui nous conduit à rejeter l'hypothèse nulle de non-stationnarité au seuil de 1%. Cela signifie que la série *dvalue* est stationnaire selon ce test. Pour compléter l'analyse, nous avons également effectué un test de Kwiatkowski–Phillips–Schmidt–Shin (KPSS), qui teste l'hypothèse nulle de stationnarité. Ce test renvoie une p-value de 0.1, donc nous ne rejetons pas l'hypothèse nulle de stationnarité au seuil de 5% ou même 1%. Cela renforce la conclusion précédente : la série différenciée *dvalue* est bien stationnaire.

1.3 Question 3

Ainsi, on retient comme série finale la série différenciée d'ordre 1. Les deux séries sont représentées ici :

FIGURE 2 – Représentation graphique de la série choisie avant et après transformation

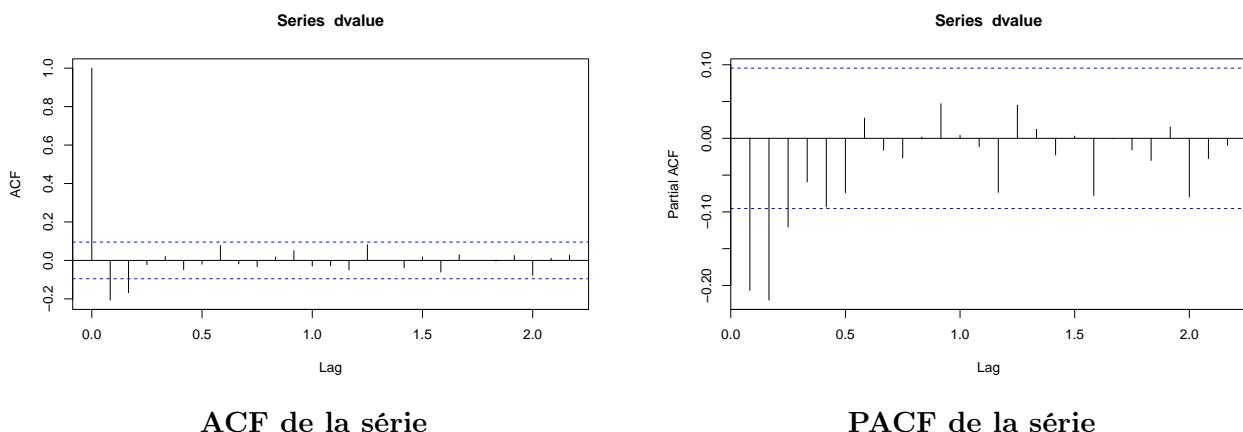


2 Partie II

2.1 Question 4

Afin de modéliser la dynamique de la série `value`, nous avons dans un premier temps centré sa première différence `dvalue`. Nous avons ensuite utilisé l'analyse des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) pour proposer des ordres initiaux de modèles ARMA 3. Ainsi, l'ACF suggère un ordre MA maximal $q = 2$, tandis que la PACF suggère un ordre AR maximal $p = 3$. Nous avons donc estimé l'ensemble des modèles ARMA(p, q) pour $p \in \{0, 1, 2, 3\}$ et $q \in \{0, 1, 2\}$, soit 12 combinaisons. Pour chaque modèle, nous avons effectué le test de Ljung-Box pour vérifier l'absence d'autocorrélation des résidus, puis testé la significativité des coefficients via des tests de Student et enfin calculé les critères d'information AIC et BIC.

FIGURE 3 – Autocorrélation (ACF) et autocorrélation partielle (PACF) de la série différenciée



Après l'estimation de tous les modèles ARMA(p, q) pour $p \in \{0, 1, 2, 3\}$ et $q \in \{0, 1, 2\}$, nous avons appliqué deux séries de tests pour évaluer la qualité de chaque ajustement :

1. Validité des résidus : Test de Ljung-Box

Le test de Ljung-Box est utilisé pour détecter la présence d'autocorrélation dans les résidus d'un modèle. Il teste l'hypothèse :

$$\begin{cases} H_0 : \text{Les résidus sont indépendants} \\ H_1 : \text{Les résidus sont autocorrélés} \end{cases}$$

Une p-value supérieure à 1 % nous empêche de rejeter H_0 à ce seuil, ce qui signifie que les résidus peuvent être considérés comme un bruit blanc, condition indispensable pour qu'un modèle ARMA soit statistiquement valide. Les modèles pour lesquels les résidus passent ce test sont les suivants :

$$\text{ARMA}(3,0), \text{ARMA}(1,1), \text{ARMA}(2,1), \text{ARMA}(3,1), \text{ARMA}(0,2), \text{ARMA}(1,2), \text{ARMA}(2,2), \\ \text{ARMA}(3,2)$$

2. Significativité des coefficients de plus grand ordre : Test de Student

Pour chaque paramètre estimé dans les modèles, un test de Student a été réalisé. L'hypothèse nulle est la suivante :

$$\begin{cases} H_0 : \text{Le coefficient est nul} \\ H_1 : \text{Le coefficient est différent de zéro} \end{cases}$$

Un coefficient est retenu s'il est significatif au seuil de 5 %, c'est-à-dire si sa p-value est inférieure à 0.05. Nous avons retenu uniquement les modèles pour lesquels **tous les coefficients** estimés sont significatifs. Les modèles satisfaisant ce critère sont :

$$\text{ARMA}(1,0), \text{ARMA}(2,0), \text{ARMA}(0,1), \text{ARMA}(1,1), \text{ARMA}(0,2)$$

3. Sélection croisée des modèles valides

En croisant les deux critères, nous retenons les modèles suivants :

$$\text{ARMA}(1,1) \quad \text{et} \quad \text{ARMA}(0,2)$$

Ce sont les deux seuls modèles qui satisfont simultanément les exigences de validité statistique des résidus et de pertinence des coefficients estimés.

4. Comparaison des modèles candidats : critères AIC, BIC et R^2 ajusté

Pour départager les modèles ARMA(1,1) et ARMA(0,2), nous avons utilisé les critères d'information classiques :

- Le **critère d'information d'Akaike (AIC)**, qui équilibre la qualité d'ajustement et la complexité du modèle. Il favorise les modèles avec le meilleur compromis entre performance et parcimonie.
- Le **critère d'information bayésien (BIC)**, plus pénalisant pour les modèles complexes. Il tend à privilégier les modèles plus simples, à condition qu'ils expliquent bien les données.

Les résultats 1 montrent que le modèle ARMA(0, 2) minimise l'AIC mais également le BIC. Enfin, un critère complémentaire a été calculé : le R^2 ajusté, évalué à partir de la part de variance expliquée par le modèle, corrigée du nombre de paramètres estimés. Ce critère confirme que le modèle ARMA(0, 2) fournit la meilleure qualité d'ajustement.

2.2 Question 5

La série `dvalue` étant la première différence de la série `value`, nous en déduisons que le modèle retenu pour la série initiale est un :

$$\boxed{\text{ARIMA}(0, 1, 2)}$$

3 Partie III

3.1 Question 6

Nous allons maintenant procéder à l'élaboration de prévisions pour la série de données sélectionnée. Nous commencerons par le modèle ARIMA(0,1,2), dont les prévisions se révèlent constantes au-delà de l'horizon 2. Nous examinerons ensuite le modèle ARIMA(1,1,1) en annexe, qui présente un intérêt accru notamment si on effectue des prévisions à horizon $h > 2$. En effet, bien que nous effectuions une prévision à horizon $h = 2$, il convient de souligner cette spécificité du modèle ARIMA(0,1,2) dans l'éventualité où une prévision à plus long terme serait envisagée, auquel cas le modèle ARIMA(1,1,1) apparaîtrait comme plus approprié.

Modèle ARIMA(0,1,2)

La formule générale du modèle ARIMA(0,1,2) est

$$(1 - L)X_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \Rightarrow X_t = X_{t-1} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$$

Pour $h > 2$:

$$X_{t,t+h} = EL[X_{t+h} | \underline{X_t}] = EL[X_{t+h-1} + \epsilon_{t+h} + \theta_1\epsilon_{t+h-1} + \theta_2\epsilon_{t+h-2} | \underline{X_t}] = E[X_{t+h-1} | \underline{X_t}] = X_{t,t+h-1}$$

Comme $(\epsilon_t)_t$ est le bruit blanc des innovations et $t+h-2 > t$, cela implique l'absence des (ϵ_{t+h}) dans l'expression finale. La prédiction est donc constante à partir de $h \geq 2$.

Pour $h = 2$:

$$X_{t,t+2} = EL[X_{t+2} | \underline{X_t}] = EL[X_{t+1} + \epsilon_{t+2} + \theta_1\epsilon_{t+1} + \theta_2\epsilon_t | \underline{X_t}] = X_{t,t+1} + \theta_2\epsilon_t$$

Pour $h = 1$:

$$X_{t,t+1} = EL[X_{t+1} | \underline{X_t}] = EL[X_t + \epsilon_{t+1} + \theta_1\epsilon_t + \theta_2\epsilon_{t-1} | \underline{X_t}] = X_t + \theta_1\epsilon_t + \theta_2\epsilon_{t-1}$$

On pose $e_h = X_t - X_{t,t+h} \in Vect(\epsilon_{t+1}, \dots, \epsilon_{t+h})$. On peut alors montrer

$$\forall h \geq 0, \quad e_h = \sum_{i=0}^{h-1} a_i \epsilon_{t+h-i}$$

où les coefficients $(a_n)_{n \in \mathbb{N}}$ sont ceux du polynôme :

$$a(L) = \frac{1 + \theta_1 L + \theta_2 L^2}{(1 - L)}$$

Après une décomposition en éléments simples on obtient :

$$a(L) = \sum_{k=0}^{\infty} a_k L^k \quad \text{avec} \quad \begin{cases} a_0 = 1 \\ a_1 = 1 + \theta_1 \\ a_k = 1 + \theta_1 + \theta_2 \quad \text{pour } k \geq 2 \end{cases}$$

Comme $e_h = \sum_{i=0}^{h-1} a_i \varepsilon_{t+h-i}$, on a : $\boxed{\text{Var}(e_h) = \sigma^2 \sum_{i=0}^{h-1} a_i^2}$

Enfin, puisque e_h est une combinaison linéaire de loi normales centrées, il vient $e_h \sim \mathcal{N}(0, \text{Var}(e_h))$. Ainsi, on en déduit,

$$\boxed{\forall h \geq 0, X_{t+h} \in [X_{t+h} - 1.96\sqrt{\text{Var}(e_h)}; X_{t+h} + 1.96\sqrt{\text{Var}(e_h)}]}$$

3.2 Question 7

Les résultats présentés concernant les intervalles de confiance à 95% reposent sur plusieurs hypothèses fondamentales, qu'il convient d'expliciter :

- Le modèle ARMA utilisé est supposé parfaitement identifié, c'est-à-dire qu'il n'y a pas de racines communes entre les polynômes AR et MA.
- Le bruit blanc d'innovation suit une loi normale centrée : $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, avec la variance $\sigma^2 > 0$ connue.

Ces hypothèses permettent de calculer analytiquement les expressions des prévisions conditionnelles $X_{t,t+1}$ et $X_{t,t+2}$, ainsi que leurs intervalles de confiance à partir de la variance des résidus.

3.3 Question 8

Dans cette section, nous présentons une représentation graphique des prédictions effectuées à l'horizon 2 pour notre série, en utilisant un modèle ARIMA(0,1,2).

FIGURE 4 – Prévision à horizon 2 d'un modèle ARIMA (0,1,2)

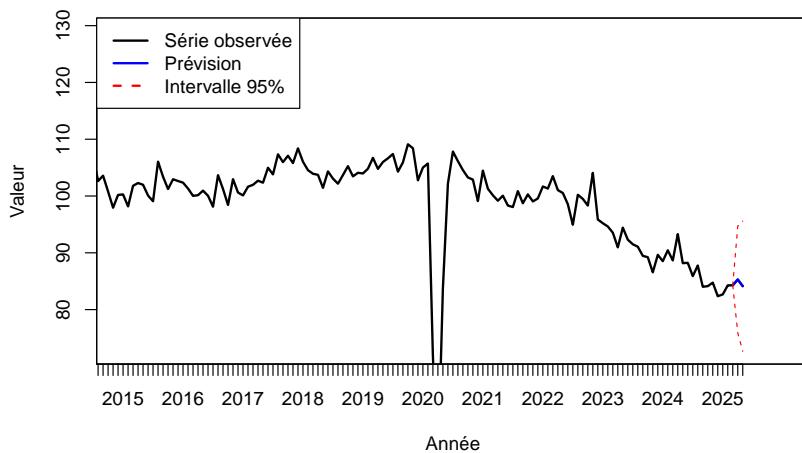
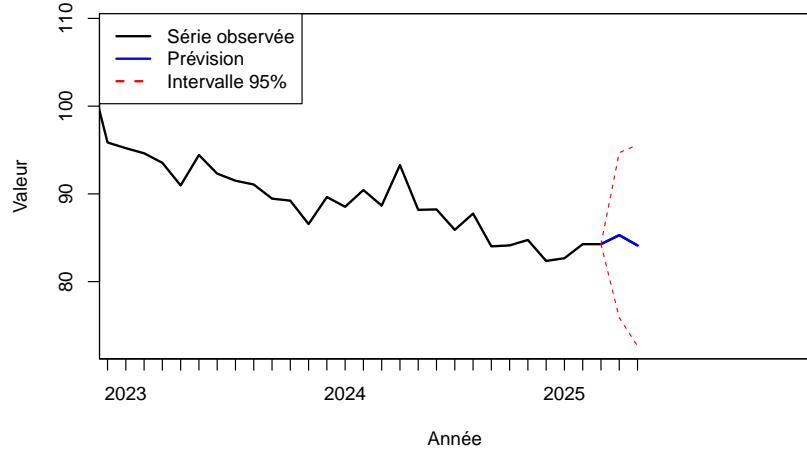


FIGURE 5 – Prévision à horizon 2 d'un modèle ARIMA (0,1,2)



3.4 Question 9

Pour que l'information de Y_{T+1} améliore la prévision de X_{T+1} il faut que la série (Y_t) cause instantanément la série (X_t) au sens de Granger. La série (Y_t) cause instantanément (X_t) si l'information contenue dans Y_{T+1} permet de réduire l'erreur quadratique moyenne de prévision de X_{T+1} , comparée à une prévision basée uniquement sur les informations passées $(X_t)_{t \leq T}$. Autrement dit, la connaissance de Y_t , même contemporaine à X_t , permet d'améliorer la prédition de X_t lorsqu'on la combine aux valeurs passées de X_t . Dans ce contexte, si Y_{t+1} est observé avant X_{t+1} , alors cette information peut effectivement être utilisée pour affiner la prévision de X_{t+1} à condition que cette causalité instantanée soit statistiquement vérifiée.

Avant de tester la causalité de Granger entre deux séries potentiellement non stationnaires, il est nécessaire de tester leur cointégration. En effet, si les séries sont I(1) mais non cointégrées, alors tout test effectué sur les séries en niveau serait invalide, car il pourrait conduire à des résultats fallacieux. Un test de Wald permet ensuite d'évaluer la significativité du coefficient des résidus de Y_t dans la régression des résidus de X_t sur les résidus de Y_t . Si ce coefficient est significativement non nul, on rejette l'hypothèse que Y_t ne cause pas instantanément X_t au sens de Granger.

4 Annexes

FIGURE 6 – Représentation de la série différenciée

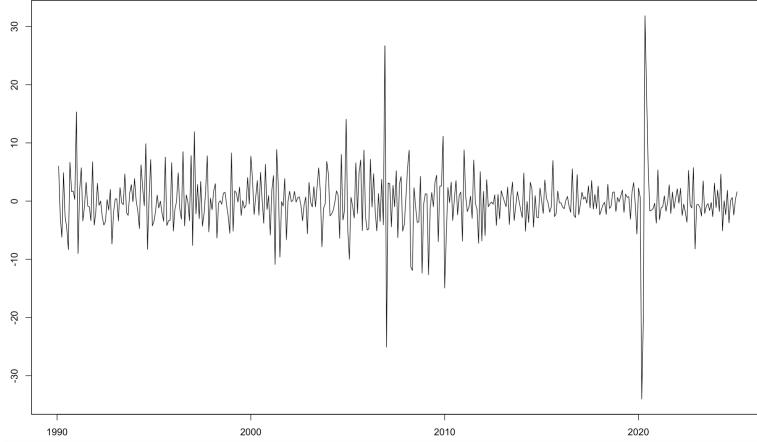


TABLE 1 – Comparaison des modèles ARMA(0,2) et ARMA(1,1) selon différents critères

Critère	ARMA(0,2)	ARMA(1,1)
AIC	2519.633	2521.283
BIC	2531.761	2533.411
R^2 ajusté	0.1094058	0.1060834

Modèle ARIMA(1,1,1)

La formule générale du modèle ARIMA(1,1,1) est

$$(1 - L)(1 - \phi L)X_t = (1 + \theta L)\epsilon_t \Rightarrow (1 - (1 + \phi)L + \phi L^2)X_t = (1 + \theta L)\epsilon_t$$

Comme précédemment, pour un horizon $h > q = 1$ et puisque (ϵ_t) est le bruit blanc des innovations on a

$$\forall h \geq 2, \quad X_{t,t+h} = (1 + \phi)X_{t,t+h-1} - \phi X_{t,t+h-2}$$

De plus, on a

$$X_{t,t} = X_t \quad \text{et} \quad X_{t,t+1} = EL[(1 + \phi)X_t - \phi X_{t-1} + \epsilon_{t+1} + \theta \epsilon_t | \underline{X}_t] = X_t + \phi(X_t - X_{t-1}) + \theta \epsilon_t$$

En résolvant cette relation de récurrence avec les conditions initiales, on aboutit à

$$\boxed{\forall h \geq 0, X_{t,t+h} = \frac{1}{1 - \phi}(X_{t,t+1} - \phi X_t + (X_t - X_{t-1})\phi^h)}$$

En effet, nous avons précédemment établi que la suite des prédictions associées à un modèle ARIMA(1,1,1) obéit à la formule de récurrence suivante :

$$\forall h \geq 2, \quad X_{t,t+h} = (1 + \phi)X_{t,t+h-1} - \phi X_{t,t+h-2}$$

accompagnée des conditions initiales :

$$X_{t,t} = X_t \quad \text{et} \quad X_{t,t+1} = \mathbb{E} [(1 + \phi) X_t - \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \mid \underline{X}_t] = X_t + \phi(X_t - X_{t-1}) + \theta \epsilon_t$$

L'équation caractéristique associée à cette récurrence linéaire d'ordre 2 s'écrit :

$$r^2 - (1 + \phi)r + \phi = 0 \text{ ce qui équivaut à } (r - 1)(r - \phi) = 0$$

Par conséquent, la solution générale de la suite $(X_{t,t+h})_{h \in \mathbb{N}}$ s'écrit sous la forme :

$$X_{t,t+h} = A \cdot 1^h + B \cdot \phi^h = A + B\phi^h$$

En utilisant les conditions initiales à notre disposition, nous obtenons le système suivant, qui permet de déterminer les constantes A et B :

$$A + B = X_t, \quad \text{et} \quad A + B\phi = X_{t,t+1}$$

La résolution de ce système nous donne alors les expressions explicites suivantes :

$$B = \frac{X_t - X_{t,t+1}}{1 - \phi}, \quad A = X_t - B = \frac{X_{t,t+1} - \phi X_t}{1 - \phi}$$

Nous obtenons ainsi la forme finale de la suite prédictive $(X_{t,t+h})_{h \in \mathbb{N}}$, exprimée comme suit :

$$\forall h \geq 0, \quad X_{t,t+h} = \frac{1}{1 - \phi} (X_{t,t+1} - \phi X_t + (X_t - X_{t,t+1})\phi^h)$$

On pose $e_h = X_t - X_{t,t+h} \in Vect(\epsilon_{t+1}, \dots, \epsilon_{t+h})$. On peut alors montrer

$$\forall h \geq 0, \quad e_h = \sum_{i=0}^{h-1} a_i \epsilon_{t+h-i}$$

où les coefficients $(a_n)_{n \in \mathbb{N}}$ sont ceux du polynôme :

$$a(L) = \frac{1 + \theta L}{(1 - L)(1 - \phi L)}$$

Une décomposition en éléments simples permet de trouver le polynôme a et il vient

$$a(L) = 1 + \sum_{k=1}^{+\infty} \frac{1 - \phi^{k+1} + \theta - \theta \phi^k}{1 - \phi} L^k$$

En effet, le polynôme a est défini par

$$a(L) = \frac{1 + \theta L}{(1 - L)(1 - \phi L)}$$

La décomposition en éléments simples du dénominateur fait apparaître deux constantes que l'on peut facilement déterminer. En effet, pour le coefficient α , il suffit de multiplier l'équation par le polynôme $P(L) = 1 - L$ et de spécialiser en 1.

$$\frac{1}{(1 - L)(1 - \phi L)} = \frac{\alpha}{1 - L} + \frac{\beta}{1 - \phi L} \Rightarrow \begin{cases} \alpha = \frac{1}{1 - \phi} \\ \beta = \frac{\phi}{\phi - 1} \end{cases}$$

On obtient alors la formule du polynôme a : $a(L) = (1+\theta L) \left(\frac{1}{(1-\phi)(1-L)} + \frac{\phi}{\phi-1} \cdot \frac{1}{1-\phi L} \right)$

On procède ensuite à un développement en séries entières afin de déterminer les inverses des polynômes du dénominateur :

$$a(L) = (1 + \theta L) \left(\frac{1}{1-\phi} \sum_{k=0}^{+\infty} L^k + \frac{\phi}{\phi-1} \sum_{k=0}^{+\infty} \phi^k L^k \right) = (1 + \theta L) \sum_{k=0}^{+\infty} \frac{1 - \phi^{k+1}}{1-\phi} L^k$$

En développant cette expression, on obtient finalement la forme suivante du polynôme $a(L)$:

$$a(L) = \sum_{k=0}^{+\infty} \frac{1 - \phi^{k+1}}{1-\phi} L^k + \theta \sum_{k=0}^{+\infty} \frac{1 - \phi^{k+1}}{1-\phi} L^{k+1} = 1 + \sum_{k=1}^{+\infty} \frac{1 - \phi^{k+1} + \theta - \theta \phi^k}{1-\phi} L^k$$

On peut dorénavant étudier la variance de ces écarts et puisque (ϵ_t) est le bruit blanc des innovations on a

$$\forall h \geq 0, \text{Var}(e_h) = \sigma^2 \sum_{k=0}^{h-1} \frac{(1 - \phi^{k+1} + \theta - \theta \phi^k)^2}{(1-\phi)^2}$$

Enfin, puisque e_h est une combinaison linéaire de loi normales centrées, il vient $e_h \sim \mathcal{N}(0, \text{Var}(e_h))$. Ainsi, on en déduit,

$$\forall h \geq 0, X_{t+h} \in [X_{t,t+h} - 1.96\sqrt{\text{Var}(e_h)}; X_{t,t+h} + 1.96\sqrt{\text{Var}(e_h)}]$$