# Séries temporelles

Vincent Lefieux

ENSAE-ENSAI
Formation continue
(Cepe)

# Plan

# Plan

Généralités

# Partition de musique

- Il est rare d'en trouver une modélisation mathématique (à l'exception de la musique sérielle de Schönberg).


Figure – Chanson patriotique, partition manuscrite de la fin du 18ème siècle (Source : Archives départementales du Pas-de-Calais)

# Bruit blanc

- Un bruit blanc gaussien est constitué de variables aléatoires i.i.d de loi normale.

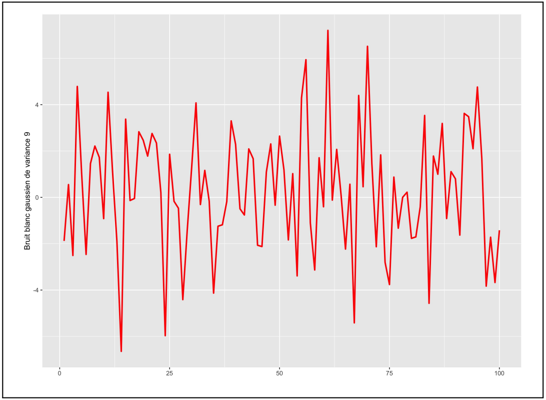


Figure – Exemple de bruit blanc gaussien de variance 9 (n=100)

# Population des Etats Unis

- La série semble présenter une tendance croissante (parabolique ? exponentielle ?).

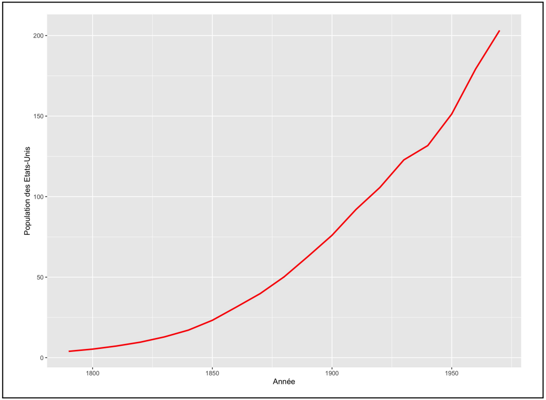


Figure – Population des Etats Unis de 1790 à 1990 (Au pas de temps décennal)

# Nombre mensuel de passagers aériens

- La série semble présenter une tendance croissante (parabolique ? exponentielle ?) et des hausses/baisses régulièrement espacées.

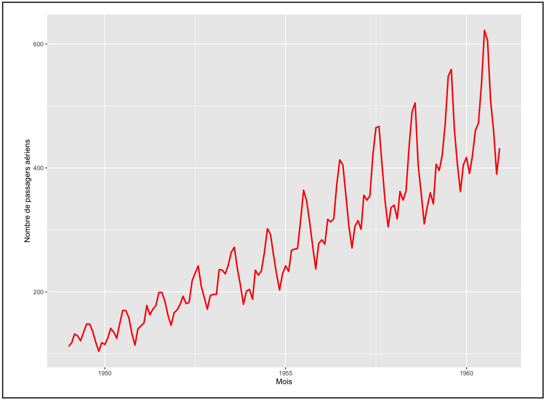


Figure – Nombre mensuel de passagers aériens sur les vols internationaux de janvier 1949 à décembre 1960

# Nombre mensuel de passagers aériens (Log)

- La série semble présenter une tendance croissante (linéaire ?) et des hausses/baisses périodiques (saisonnalité de période 12).

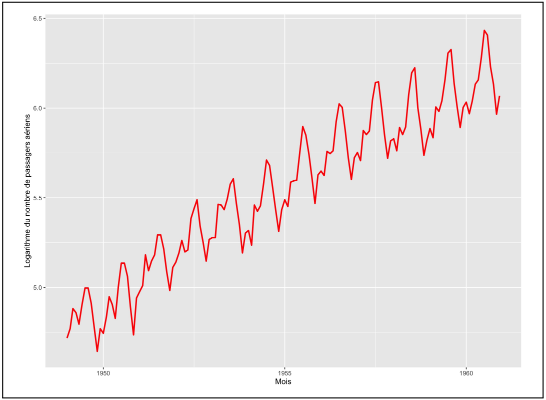


Figure – Logarithme du nombre mensuel de passagers aériens sur les vols internationaux de janvier 1949 à décembre 1960

# Production mensuelle de bière australienne

- La série semble présenter une tendance croissante puis stable (avec quelques décroissances) et une saisonnalité de période 12.

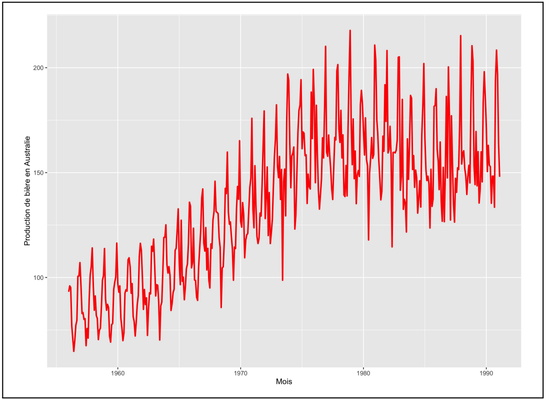


Figure – Production mensuelle de bière en Australie entre janvier 1956 et février 1991

# Nombre annuel de lynx capturés au Canada

- La série semble présenter une saisonnalité de période 10 et une autre de période 40 (cycle).

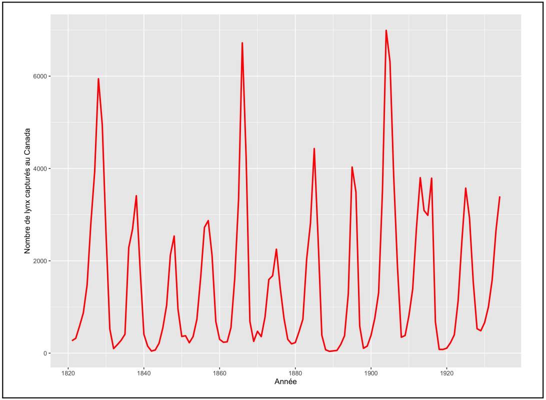


Figure – Nombre annuel de lynx capturés au Canada entre 1821 et 1934

# Notion de stationnarité



- Dans de très nombreux cas, on ne peut pas renouveler la suite de mesures dans des conditions identiques.
- Pour que le modèle déduit à partir d'une suite d'observations ait un sens, il faut que toute portion de la trajectoire observée fournisse des informations sur la loi du processus et que des portions différentes, mais de même longueur, fournissent les mêmes indications.
- D'où la notion de stationnarité.

# Processus (faiblement) stationnaire



- Un processus $(X_t)_{t\in\mathbb{Z}}$ du second ordre est faiblement stationnaire si son espérance et ses autocovariances sont invariantes par translation dans le temps :
  - $\forall t \in \mathbb{Z} : \mathbb{E}(X_t) = \mu$ ,
  - $\forall (s,t) \in \mathbb{Z}^2, \forall h \in \mathbb{Z} : \mathrm{Cov}(X_s, X_t) = \mathrm{Cov}(X_{s+h}, X_{t+h})$.

# Traiter la non-stationnarité I



- $(X_t)_{t\in\mathbb{Z}}$ est un processus non-stationnaire TS (Trend Stationary) s'il peut s'écrire sous la forme :

  $$X_t = f(t) + Y_t$$

  où $f$ est une fonction déterministe et $(Y_t)_{t\in\mathbb{Z}}$ un processus stationnaire.

- La non-stationnarité TS peut se traiter à l'aide de la décomposition saisonnière.

# Traiter la non-stationnarité II

- On considère l'opérateur retard :

$$BX_t = X_{t-1}$$

ainsi que l'opérateur différence :

$$\nabla^i X_t = (I - B)^i X_t \ .$$

- $(X_t)_{t\in\mathbb{Z}}$ est un processus non-stationnaire DS (Difference Stationary) si :
    - $\left(\nabla^i X_t\right)_{t\in\mathbb{Z}}$, $i \in \{0, \ldots, d-1\}$, n'est pas stationnaire,
    - $\left(\nabla^d X_t\right)_{t\in\mathbb{Z}}$ est stationnaire.

  Un tel processus est rendu stationnaire après $d$ différenciations.

# Fonction d'autocovariance



- On appelle fonction d'autocovariance d'un processus $(X_t)_{t\in\mathbb{Z}}$ stationnaire la fonction $\gamma$ suivante :

$$\forall h \in \mathbb{Z} : \gamma(h) = \mathrm{Cov}\left(X_t, X_{t-h}\right) .$$

- $\gamma$ est une fonction symétrique :

$$\forall h \in \mathbb{Z} : \gamma(-h) = \gamma(h) .$$

On pourra donc calculer la fonction d'autocovariance pour $h \in \mathbb{N}$ .

# Autocorrélogramme simple



- On appelle autocorrélogramme simple d'un processus $(X_t)_{t\in\mathbb{Z}}$ stationnaire la fonction $\rho$ suivante :

$$\forall h \in \mathbb{Z} : \rho(h) = \mathrm{Corr}(X_t, X_{t-h}) = \frac{\gamma(h)}{\gamma(0)} \ .$$

- On a $|\rho(h)| \leq \rho(0) = 1$

# Autocorrélogramme partiel



- On appelle autocorrélogramme partiel d'un processus $(X_t)_{t\in\mathbb{Z}}$ stationnaire la fonction $r$ suivante :
  - $r(0) = 1$ .
  - $r(1) = \rho(1)$ .
  - $\forall h \in \mathbb{N} \setminus \{0, 1\}$ :

$$r(h) = \mathrm{Corr}\left(X_t, X_{t-h} / X_{t-1}, \ldots, X_{t-h+1}\right) .$$

# Estimation des moments pour les processus stationnaires : espérance



- L'estimateur naturel (consistant et sans biais) de $\mathbb{E}\left(X_t\right) = \mu$ à partir de $(X_1, \ldots, X_T)$ est la moyenne empirique $\overline{X}_T$ :
$$\widehat{\mu} = \overline{X}_T$$
avec $\overline{X}_T = \frac{1}{T}\sum_{i=1}^{T} X_i$ .

# Estimation des moments pour les processus stationnaires : autocorrélations



- A partir de $(X_1, \ldots, X_T)$, on peut considérer les estimateurs suivants :

$$\forall h \in \mathbb{N}^\star : \widehat{\gamma}(h) = \frac{1}{T-h} \sum_{t=h+1}^{T} \left(X_t - \overline{X}_T\right)\left(X_{t-h} - \overline{X}_T\right).$$

- On en déduit :

$$\forall h \geq 1 : \widehat{\rho}(h) = \frac{T}{T-h} \frac{\sum_{t=h+1}^{T} \left(X_t - \overline{X}_T\right)\left(X_{t-h} - \overline{X}_T\right)}{\sum_{t=1}^{T} \left(X_t - \overline{X}_T\right)^2}.$$

- Les estimations des autocorrélations partielles se déduisent des estimations des autocorrélations simples grâce à l'algorithme de Durbin-Levinson.

# Test de Portmanteau I

- On considère le test suivant pour un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ :

$$\begin{cases} H_0 : \rho(1) = \ldots = \rho(k) = 0 \\ H_1 : H_0 \text{ est fausse} \end{cases} .$$

- Si on dispose de $(X_1, \ldots, X_T)$, on considère la statistique de Portmanteau (calculée sur les $k$ premières estimations des autocorrélations) :

$$Q_k = T \sum_{h=1}^{k} \widehat{\rho}^2(h) .$$

- Une trop grande valeur de $Q_k$ indique que les autocorrélations sont trop importantes pour être celles d'un bruit blanc.

# Test de Portmanteau II

- Il existe d'autres versions de cette statistiques, par exemple la statistique de Ljung–Box :

$$Q_k^* = T(T+2)\sum_{h=1}^{k} \frac{\widehat{\rho}^2(h)}{T-h} \ .$$

- Sous $H_0$, $Q_k$ (ainsi que $Q_k^*$) suit asymptotiquement une loi du Khi-deux à $k$ degrés de liberté :

$$Q_k \xrightarrow{\mathcal{L}} \chi^2(k) \ .$$

- On rejette donc l'hypothèse $H_0$ au niveau de test $\alpha$ si :

$$Q_k > \chi^2_{k,1-\alpha}$$

  où $\chi^2_{k,1-\alpha}$ désigne le quantile d'ordre $1-\alpha$ d'une loi du Khi-deux à $k$ degrés de liberté.

- La p-valeur vaut :

$$\text{p-valeur} = \mathbb{P}\left(\chi^2_k \geq Q_k\right) \ .$$

# Pourquoi les modèles ARMA ?



- La décomposition de Wold nous indique que tout processus stationnaire peut être modélisé par un processus ARMA.

# Plan

# Plan

# Processus AR



- Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc faible de variance $\sigma^2$.
- On dit qu'un processus $(X_t)_{t\in\mathbb{Z}}$ est un processus processus AR (AutoRegressive) d'ordre $p$ si :
  - $(X_t)_{t\in\mathbb{Z}}$ est stationnaire,
  - $\forall t \in \mathbb{Z} : X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$ .
- On utilise la notation $\mathrm{AR}(p)$.

# Autocorrélations simples d'un processus AR



- Les autocorrélations simples décroissent vers 0, de manière exponentielle ou sinusoïdale amortie.

# Autocorrélations partielles d'un processus AR



- Si $(X_t)_{t\in\mathbb{Z}}$ est un processus $\text{AR}(p)$ alors ses autocorrélations partielles s'annulent à partir du rang $p+1$ :

$$\begin{cases} r(p) \neq 0 \\ \forall h \in \mathbb{N}, h \geq p+1 : r(h) = 0 \end{cases}.$$

- Réciproquement, il s'agit d'une condition nécessaire et suffisante pour qu'un processus $(X_t)_{t\in\mathbb{Z}}$ soit un $\text{AR}(p)$ .

# Processus MA



- Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc faible de variance $\sigma^2$.
- On dit qu'un processus $(X_t)_{t\in\mathbb{Z}}$ est un processus MA (Moving Average) d'ordre $q$ si :

$$\forall t \in \mathbb{Z} : X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \ .$$

- On utilise la notation $\mathrm{MA}(q)$.

# Autocorrélations simples d'un processus MA

- Si $(X_t)_{t\in\mathbb{Z}}$ est un processus MA $(q)$ alors ses autocorrélations simples s'annulent à partir du rang $q+1$ :

$$\begin{cases} \rho\left(q\right) \neq 0 \\ \forall h \in \mathbb{N}, h \geq q+1 : \rho\left(h\right) = 0 \end{cases} .$$

- Il s'agit d'une condition nécessaire et suffisante pour qu'un processus $(X_t)_{t\in\mathbb{Z}}$ soit un MA $(q)$.

- Il n'y a rien de notable pour les autocorrélations partielles d'un processus MA.

# Processus ARMA



- Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc faible de variance $\sigma^2$.
- On dit qu'un processus $(X_t)_{t\in\mathbb{Z}}$ est un processus ARMA (AutoRegressive Moving Average) d'ordre $(p, q)$ si :
  - $(X_t)_{t\in\mathbb{Z}}$ est stationnaire,
  - $\forall t \in \mathbb{Z} : X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$ .
- On utilise la notation ARMA $(p, q)$.

# Remarques

- Un processus $\mathrm{AR}(p)$ est un processus $\mathrm{ARMA}(p, 0)$.

- Un processus $\mathrm{MA}(q)$ est un processus $\mathrm{ARMA}(0, q)$.

# Autocorrélations simples d'un processus ARMA



- Les autocorrélations simples décroissent vers 0.
- Au-delà de ce résultat, il n'existe pas de règle simple pour lier les autocorrélations (simples et partielles) et les ordres d'un processus ARMA.

# Estimation des processus ARMA



- Afin d'estimer un processus ARMA, on utilise généralement un algorithme de type maximum de vraisemblance (en considérant une distribution gaussienne), avec une initialisation basée sur les autocorrélations.

# Processus ARIMA

- Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc faible de variance $\sigma^2$.
- On dit qu'un processus $(X_t)_{t\in\mathbb{N}}$ est un processus ARIMA (AutoRegressive Integrated Moving Average) d'ordre $(p, d, q)$ si :

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

  où :
  - $\nabla^d = (I - B)^d$ ,
  - $\Phi(B) = I - \varphi_1 B - \ldots - \varphi_p B^p$ ,
  - $\Theta(B) = I + \theta_1 B + \ldots + \theta_q B^q$ .
- On utilise la notation $\text{ARIMA}(p, d, q)$.

# Remarques



- Les modèles ARIMA permettent de modéliser des séries temporelles qui présentent une tendance polynomiale.
- $(I - B)^d X_t$ est équivalent asymptotiquement à un processus ARMA $(p, q)$.

# Processus SARIMA

- Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc faible de variance $\sigma^2$.
- On dit qu'un processus $(X_t)_{t\in\mathbb{N}}$ est un processus SARIMA (Seasonnal AutoRegressive Integrated Moving Average) d'ordre $(p,d,q)\,(P,D,Q)_s$ si :

$$\Phi\left(B\right)\Phi'\left(B^s\right)\nabla^d\nabla_s^d X_t = \Theta\left(B\right)\Theta'\left(B^s\right)\varepsilon_t$$

  où :
  - $\nabla^d = \left(I - B\right)^d$ ,
  - $\nabla_s^d = \left(I - B^s\right)^d$ ,
  - $\Phi\left(B\right) = I - \varphi_1 B - \ldots - \varphi_p B^p$ ,
  - $\Phi'\left(B\right) = I - \varphi'_1 B - \ldots - \varphi'_P B^P$ ,
  - $\Theta\left(B\right) = I + \theta_1 B + \ldots + \theta_q B^q$ ,
  - $\Theta'\left(B\right) = I + \theta'_1 B + \ldots + \theta'_Q B^Q$ .
- On utilise la notation $\mathrm{SARIMA}\,(p,d,q)\,(P,D,Q)_s$.

# Remarque

- Les modèles SARIMA permettent de modéliser des séries qui présentent une saisonnalité.
- Estimer un modèle SARIMA se ramène en pratique à l'estimation d'un modèle ARMA sur la série différenciée.

# Plan

Modèles ARMA (et SARIMA

Définitions et propriétés

Pratique

# Vue synthétique de la démarche

1. Stationnarisation (éventuellement)
2. Identification a priori de modèles potentiels
3. Estimation des modèles potentiels
4. Vérification des modèles potentiels
5. Choix définitif d'un modèle
6. Prévision à l'aide du modèle choisi
7. Analyse a posteriori de la prévision

# Stationnarisation

- La plupart des séries temporelles présentent une tendance et/ou une saisonnalité, et ne sont donc pas modélisables par un processus stationnaire.
- Décomposition saisonnière
  Cette méthode permet d'éliminer tendance et saisonnalité, sources évidentes de non-stationnarité. Il se peut néanmoins que la série résultant de la décomposition ne soit toujours pas stationnaire.
- Différenciation
  C'est la méthode employée par les modèles ARIMA et SARIMA. On ne modélise pas la série brute mais la série différenciée, en « tendance » à l'aide de $\nabla^d$, ou la série différenciée en « saisonnalité » à l'aide de $\nabla_s^d$. De manière générale, $\nabla^d$ permet de stationnariser des séries possédant une tendance polynomiale de degré $d$, et $\nabla_s^d$ des séries possédant une composante saisonnière de période $s$.

# Remarque

- Attention, si les processus stationnaires peuvent être approchés par des modèles ARMA, il n'est pas certain que la différenciation permette de stationnariser tous les processus.

# Méthode empirique

- Dans le cas des modèles SARIMA, on utilise très souvent une méthode empirique basée sur l'autocorrélogramme simple.
- On effectue une différenciation en « tendance » si :
  - Les autocorrélations « $ACF(h)$ » sont proches de 1 pour un grand nombre de retards.
  - Les premières autocorrélations « $ACF(h)$ » sont proches les unes des autres (même si elles ne sont pas forcément proches de 1).

  On parle souvent de décroissance lente des autocorrélations simples.
- On effectue une différenciation en « saisonnalité » si des comportements similaires sont observés de manière périodique. Par exemple, si $ACF(12)$, $ACF(24)$... sont proches de 1, on utilise alors une différenciation en « saisonnalité » avec $s = 12$.

# Remarques



- On peut compléter cette méthode empirique par des tests de racines unité qui permettent de détecter des problèmes de stationnarité pour des séries qui ne présentent pas forcément de tendance et/ou de saisonnalité (non-stationnarité dite « stochastique »).
- Quelle(s) que soi(en)t la(les) méthode(s) retenue(s), on procède de manière itérative : on effectue une première différenciation, si celle-ci d'avère insuffisante, on en effectue une seconde, etc.
- En pratique on considère rarement $d > 2$ et $D > 2$.

# Identification a priori de modèles potentiels

- Critère d'information
  Les algorithmes de recherche automatique recherchent les ordres du modèles qui minimisent un critère d'information, avec plus ou moins de succès. . .

- Heuristique
  En pratique (surtout pour les modèles SARIMA), on essaye d'identifier les autocorrélations simples et partielles « significatives » pour caler ensuite des polynômes AR et MA qui reflètent ces liens temporels. Afin d'obtenir des modèles potentiels, l'idéal est de regarder l'autocorrélogramme partiel afin d'émettre une hypothèse sur la partie autorégressive (simple et saisonnière), la tester puis regarder l'autocorrélogramme simple (et partiel) du résidu afin d'identifier complètement un modèle. Cette démarche itérative permet en général d'obtenir plusieurs modèles potentiels.

# Estimation des modèles potentiels



- On estime les modèles potentiels à l'aide des méthodes classiques : maximum de vraisemblance ou moindres carrés.
- L'estimation est effectuée sur la série préalablement différenciée.

# Vérification des modèles potentiels : significativité des paramètres

- Par exemple, pour le coefficient AR d'ordre $p$, on effectue le test suivant :

$$\begin{cases} H_0 : \text{le processus est un ARMA}(p-1,q) & (\varphi_p = 0) \\ H_1 : \text{le processus est un ARMA}(p,q) & (\varphi_p \neq 0) \end{cases}.$$

- On utilise pour cela la statistique de test suivante :

$$T = \frac{|\widehat{\varphi}_p|}{\sqrt{\widehat{\text{Var}}(\widehat{\varphi}_p)}}.$$

- Le test de Student permet de rejeter $H_0$ au niveau 5% si $|t|$ est supérieur à 1.96.
- Il existe des résultats similaire pour les coefficients MA.

# Vérification des modèles potentiels : blancheur et normalité du résidu



- On vérifie que le résidu est bien un bruit blanc, à l'aide du test de Ljung–Box par exemple, et qu'il est distribué suivant une loi normale à l'aide du test de Shapiro-Wilk par exemple.
- Remarquons que si le test de normalité n'est pas vérifié, il est utile d'étudier le caractère ARCH ou GARCH du résidu.

# Choix définitif d'un modèle

- Ce choix s'opère entre les modèles potentiels retenus via :
  - Des critères d'information basés sur l'information de Kullback, comme :
    - Le critère d'Akaïke :

$$\text{AIC}\,(p,q) = \ln\left(\widehat{\sigma}^2\right) + 2\,\frac{p+q}{T}\ .$$

    - Le critère de Schwarz :

$$\text{BIC}\,(p,q) = \ln\left(\widehat{\sigma}^2\right) + (p+q)\,\frac{\ln\left(T\right)}{T}\ .$$

  - Des critères basés sur le pouvoir prédictif.
- Une fois ce choix effectué, le modèle retenu est utilisé à des fins de prévision.

# Analyse a posteriori



- L'analyse a posteriori permet de quantifier les écarts entre les prévisions et les réalisations, en tronquant la série d'un certain nombre de points (notons que le modèle doit être correctement estimé sur la série tronquée).

- On utilise des critères d'erreur comme l'erreur quadratique moyenne (Root Mean Square Error : RMSE) ou l'erreur relative absolue moyenne (Mean Average Percentage Error : MAPE) :

$$\text{RMSE}= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \widehat{x}_t)^2} \;,$$

$$\text{MAPE}= \frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_t - \widehat{x}_t}{x_t}\right| \;.$$

# Références I