

# Chapitre 3

## Approche de prévision développée

### Sommaire

---

3.1 Généralité .....	31
3.2 Régression linéaire par la méthode des moindres carrés .....	33
3.2.1 Principe de la régression linéaire .....	33
3.2.2 Moindres carrés généralisés .....	35
3.2.3 Significativité de la régression .....	36
3.2.4 Comparaison de modèles et critères d'information .....	37
3.2.5 Intervalle de confiance .....	38
3.2.6 Prédiction .....	39
3.3 Régression non linéaire .....	39
3.4 Prévision par le modèle ARMAX .....	40
3.4.1 Identification des résidus par la méthode des moindres carrés ordinaires MCO .....	40
3.4.2 Estimation du modèle ARMAX .....	41
3.4.3 Validation du modèle ARMAX .....	42
3.4.4 Prévision du modèle ARMAX .....	42

---

### 3.1 Généralité

Nous examinons dans ce chapitre la notion de prévision, la nécessité de prévoir la production d'énergie photovoltaïque ainsi que deux approches de prévisions proposées pour ce type d'énergie. Ce chapitre traite également les notions essentielles de la régression et discute en détail la méthode des moindres carrés ainsi que le modèle ARMAX et traite ces propriétés en tant que méthode de prévision .

Prévoir, c'est observer un ensemble de données qui permet d'envisager une situation future et d'entreprendre des actions pour y parer concrètement. Autrement dit c'est porter un jugement sur les événements ou évolutions possibles à venir en utilisant comme outils le passé et le présent. Il en résulte que les prévisions sont toujours entachées d'erreur et qu'il est possible d'en établir plusieurs pour un même événement à venir (qui constitue l'objet de la prévision). Des outils d'analyse doivent donc être développés afin de comparer et de hiérarchiser les prévisions pour discerner ce qui fait qu'on puisse, ou non, en qualifier certaines de " bonnes ".

La première approche de la prévision consiste à en mesurer les spécificités. Pour envisager une typologie des problèmes de prévision en termes d'horizon, de type de produit ou de secteur, ou en termes de but opérationnel, il est utile de dégager quelques traits qui différencient fondamentalement :

- le secteur d'activité.
- l'utilisation opérationnelle.
- la (ou les) fonction(s) utilisatrice(s) de la prévision.
- l'horizon.

L'approche est très dépendante du secteur d'activité : on ne prévoit pas des livraisons de ciment par les mêmes méthodes que des ventes de savons. Les causalités économiques sous-jacentes sont différentes suivant que le secteur est plus ou moins en amont dans le circuit industriel donc plus ou moins proche de la demande finale, suivant que le produit est stockable ou non, qu'il donne lieu à un marché de renouvellement (télévision) ou non (acier), que le produit est standardisé ou non.

### Les méthodes de prévisions

Il existe en général deux méthodes de prévision :

-Premièrement il y a *les méthodes d'extrapolation* qui utilisent le passé de la variable elle-même. Seul le passé de la variable est utilisé en vue de la prévoir sans apport d'information extérieure, nous citons à titre d'exemple : le lissage par les moyennes mobiles, la modélisation par ARMA.

-Deuxièmement entrent en jeu *les méthodes explicatives* qui à son tour utilisent les valeurs passées et présentes d'une ou de plusieurs variables pour prévoir. L'ensemble d'information utilisé comporte des facteurs extérieurs qui peuvent influencer le futur en plus du passé de la variable elle-même. Parmi ces méthodes on peut citer : la régression linéaire, modèle ARMAX.

### La prévision à court terme

Dans le cadre de ce mémoire, notre approche consiste à établir des modèles de prévision à court terme. Fondamentalement, l' horizon d'une prévision à court terme dépend du contexte étudié : en gestion ou marketing, le court terme est de l'ordre de quelques mois alors que dans le domaine de la météo, le court terme est de l'ordre d'une ou deux journée(s). Avec un horizon à court terme, on suppose que le phénomène dépend de ses valeurs passées et présentes.

### Critères de validité de la méthode de prévision

La prévision des valeurs futures d'une variable  $y$  peut se faire en utilisant différentes méthodes. Les prévisions obtenues peuvent être comparées et appréciées selon plusieurs critères parmi lesquels :

*L'erreur moyenne* (mean error) :

$$\bar{e} = \frac{1}{n} \quad (3.1)$$

Le carré moyen des erreurs (mean square error) :

$$MSE = \frac{1}{n} \sum e_t^2 \quad (3.2)$$

L'erreur quadratique moyenne (Le root mean square error) :

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{\frac{1}{n} \sum e_t^2} \quad (3.3)$$

MSE étant l'erreur de prévision pour un instant  $t$  (la différence entre la valeur observée et la valeur prévue par une méthode quelconque). La meilleure méthode est celle qui fournit les valeurs les plus faibles pour ces critères.

### Prévision de l'énergie photovoltaïque

À titre récapitulatif, les méthodes de prévision de production d'énergie photovoltaïque peuvent être classifiées en quatre approches, à savoir, approche statistique, approche d'intelligence artificielle (IA), approche physique et approche hybride.

Dans tout ce qui suit, nous nous focalisons sur la première approche. Elle repose sur des modèles statistiques de séries temporelles comme les modèles ARMA.

Or, ces derniers ne peuvent pas prendre en compte l'information climatique (les données météorologiques comme l'irradiation), qui est une information primordiale pour améliorer les prévisions photovoltaïques. Pour cela, nous avons proposé un autre type de modèle : le modèle ARMAX.

Ce modèle considère les données météorologiques, et plus particulièrement *le rayonnement solaire*, comme variables exogènes. Signalons également que des techniques avancées d'intelligence artificielle, tels que les réseaux de neurones artificiels (RNA) sont aussi une des approches statistiques intéressantes mais son étude détaillée dépasse le cadre du présent mémoire.

## 3.2 Régression linéaire par la méthode des moindres carrés

Un modèle de régression linéaire est un modèle de régression d'une variable expliquée sur une ou plusieurs variables explicatives dans laquelle on fait l'hypothèse que la fonction qui relie les variables explicatives aux variables expliquées est linéaire.

### 3.2.1 Principe de la régression linéaire

La régression linéaire décompose une variable aléatoire  $y$  en son espérance (ou moyenne), exprimé linéairement en fonction d'autres variables non aléatoire :  $1, x_2, x_3, \dots, x_k$ , plus une erreur aléatoire.  $K$  est donc le nombre de variables explicatives (ou co-variables). Par exemple, si  $K = 3$  nous avons le modèle linéaire suivant :

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad t = 1, 2, \dots, T \quad (3.4)$$

Les erreurs  $u_t$  doivent vérifier un certain nombre de présupposés :

- P1** être d'espérance nulle,  $E(u_t) = 0$  ;
- P2** avoir la même variance pour tout  $t$ ,  $var(u_t) = \sigma^2$  ;
- P3** être non corrélées entre elles,  $corr(u_t, u_s) = 0, t \neq s$  ;
- P4** être normalement distribuées.

Sous ces hypothèses, l'espérance de  $y_t$  est  $\beta_1 + \beta_2 * x_{2t} + \beta_3 * x_{3t}$ .

Nous pouvons aussi écrire matriciellement le modèle et les présupposés :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \quad X = \begin{bmatrix} 1 & x_{21} & x_{31} \\ 1 & x_{22} & x_{32} \\ \vdots & \ddots & \vdots \\ 1 & x_{2T} & x_{3T} \end{bmatrix} \quad U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{bmatrix} \quad (3.5)$$

Avec ces notations, l'équation (3.4) s'écrit comme suit

$$Y_{T*1} = X_{K*T} \beta_{T*1} + U_{T*1} \quad (3.6)$$

Et les présupposés s'expriment :

$$\begin{aligned} \mathbf{P1} : & \quad E(Y) = X\beta, \\ \mathbf{P2+P3+P4} : & \quad U \sim N(O, \sigma_u^2 I_T) \end{aligned}$$

$I_t$  est la matrice identité d'ordre T. On peut encore les formuler comme suit

$$Y \sim N(X\beta, \sigma_u^2 I_T) \quad (3.7)$$

La méthode de MCO estime  $\beta$  par la valeur qui minimise la quantité

$$(Y - X * \beta)'(Y - X * \beta)$$

dans laquelle  $X'$  est la transposée de X.

Le minimum est atteint pour

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad (3.8)$$

Si les présupposés P1 + P2 + P3 sont vérifiés, cet estimateur est sans biais et de variance minimale (théorème de Gauss-Markov). Il est normalement distribué et de matrice de covariance  $(\sigma_u^2(X'X)^{-1})$

$$\hat{\beta} \sim N(\beta, \sigma_u^2(X'X)^{-1}) \quad (3.9)$$

Les vecteurs des valeurs ajustées et des résidus sont respectivement définis par

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_T \end{bmatrix} = X\hat{\beta} \quad \text{et} \quad \hat{U} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_T \end{bmatrix} \quad (3.10)$$

La variance de l'erreur est estimée sans biais par

$$S_u^2 = \frac{1}{T-K} \sum_t (y_t - \hat{y}_t)^2 = \frac{1}{T-K} \sum_t \hat{u}_t^2 \quad (3.11)$$

On estime la matrice de covariances des paramètres en remplaçant dans  $(\sigma_u^2(X'X)^{-1})$  la variance inconnue  $(\sigma_u^2)$  par son estimateur  $S_u^2$ . On note  $S_u^2(\hat{\beta}_i)$ , l'estimation de la variance de  $\hat{\beta}_i$  ainsi obtenue, c'est l'élément  $(i, i)$  de la matrice  $(S_u^2(X'X)^{-1})$ .

Rappelons que le coefficient  $\hat{\beta}_i$  est la quantité dont augmente  $y$  quand la variable  $x_i$  augmente de 1, toutes choses égales par ailleurs; dans certains milieux scientifiques, on appelle d'ailleurs  $\hat{\beta}_i$ , *coefficient de régression partielle* de  $x_i$ . On rencontrera ce terme dans les séries temporelles quand on étudiera la régression d'une série sur son passé. Une fois la régression effectuée, il faut s'assurer que les présupposés sur les erreurs sont vérifiés. Cette vérification s'effectue sur les résidus.

Le diagramme de dispersion des résidus ou des résidus normalisés contre les valeurs ajustées permet de voir si la distribution de ces résidus est bien indépendante de la valeur ajustée. Le **QQ-plot** de normalité des résidus permet de vérifier l'hypothèse de normalité. Le logiciel R offre de nombreux diagnostics, en particulier dans **lm()**, fonction de base de la régression linéaire.

Dans le cas que nous traiterons, P4, la normalité sera systématiquement examinée; P3, la non-corrélation des erreurs sera souvent rejetée. Il faudra donc comprendre le mécanisme de cette corrélation pour en tenir compte. Une fois la corrélation modélisée, la méthode des moindres carrés généralisés (MCG) permet de l'intégrer dans l'estimation de la régression. La connaissance du mécanisme de la corrélation des erreurs au cours du temps permet quant à elle de mieux prédire la série.

### 3.2.2 Moindres carrés généralisés

Dans la régression par MCO d'une série temporelle, on constate souvent que les résidus  $\hat{u}_t$  sont auto-corrélés (si l'on calcule le coefficient de corrélation entre la série des résidus et la série des résidus retardés d'un instant, on obtient une valeur élevée). C'est un signe que les  $u_t$  le sont eux-mêmes, donc P3 ne tient pas. Dans d'autres cas, le chronogramme de la série  $y_t$  montre que les  $y_t$  ont une variance non constante (cas d'hétéroscédasticité), donc P2 ne tient pas. Si les erreurs sont normalement distribuées  $N(0, \Omega)$ , le modèle s'exprime sous forme devient en notation matricielle

$$Y \sim N(X\beta, \Omega) \quad (3.12)$$

dans laquelle  $\Omega$  est la matrice des covariances de l'erreur.

La méthode des MCG fournit l'estimation de  $\beta$  suivante

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}x'\Omega^{-1}Y \quad \tilde{\beta} \sim N(\beta, X'\Omega^{-1}X)^{-1} \quad (3.13)$$

On ne connaît pas généralement la matrice  $\Omega$ , mais l'étude des résidus de l'ajustement par MCO permet d'en découvrir la structure. Dans ce travail, les résidus de la méthode par MCO montrent une dynamique dont on s'inspire pour définir la matrice  $\Omega$ .

Même si les présupposés sont vérifiés, il n'est pas sûr qu'une régression linéaire sur un certain ensemble de données soit pertinente et significative. Heureusement, on dispose d'outils adéquats pour apprécier l'intérêt d'une telle régression.

### 3.2.3 Significativité de la régression

Le coefficient de détermination  $R^2$ , noté dans les sorties **Multiple R-squared**, est défini de façon unique quand il y a une constante dans la régression. Il est donné par

$$R^2 = \frac{\sum_{i=1}^T (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^T (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^T (\hat{u}_i)^2}{\sum_{i=1}^T (\hat{y}_i - \bar{y})^2} \quad (3.14)$$

C'est le rapport de la variabilité expliquée par la régression :  $\sum_{i=1}^T (\hat{y}_i - \bar{y})^2$  sur la variabilité totale des  $y_i = \sum_{i=1}^T (y_i - \bar{y})^2$ .

Le  $R^2$  augmente avec le nombre de variables explicatives. Ce n'est donc pas un indicateur pertinent de significativité de la régression. Il n'est intéressant que pour comparer des régressions portant sur un même nombre d'observations et de variables.

Le coefficient de détermination ajusté, **Adjusted R-squared**, est

$$R_{adj}^2 = 1 - (1 - R^2) \frac{T - 1}{T - k} \quad (3.15)$$

Pour  $T$  donné, ce coefficient est une fonction décroissante de  $k$ , qui est le nombre de variables explicatives. L'effet de l'augmentation du nombre de variables sur le  $R^2$  est pénalisant pour la valeur de ce coefficient. On peut calculer un  $R^2$  pour tout ajustement d'un modèle linéaire de l'espérance, qu'il soit obtenu par MCO ou par une autre méthode. Les deux expressions de  $R^2$  ne coïncident généralement pas en dehors de la méthode par MCO avec constante dans la régression. On estime l'espérance de  $y$  par MCG. La variabilité expliquée par la régression peut se calculer par

$$\sum_{i=1}^T (\tilde{y}_i - \bar{\tilde{y}})^2 \quad (3.16)$$

où  $\hat{y}_i$  désigne l'estimation de ces valeurs ajustées.

#### Significativité globale

La régression est dite significative, si au moins un des coefficients des variables autres que la constante est non nul. L'hypothèse nulle est : " la régression n'est pas significative ", c'est-à-dire :

$H_0 : \beta_2 = \dots = \beta_k = 0$  contre  $H_1 : \text{au moins un de ces coefficients est non nul.}$

#### Statistique de test

$H_0$  exprime une contrainte sur les paramètres. Notons  $SCR_{libre}$  et  $SCR_{contrainte}$  les sommes des carrés des résidus dans l'estimation libre et dans l'estimation contrainte, c'est-à-dire sous  $H_0$ , on a

$$SCR_{libre} = \sum_t (y_t - \bar{y})^2 \text{ et } SCR_{contrainte} = \sum_t (y_t - \hat{y}_t)^2.$$

La statistique pour tester  $H_0$  contre  $H_1$  est :

$$F = \frac{SCR_{contrainte} - SCR_{libre}/(k-1)}{SCR_{libre}/(T-k)} \quad (3.17)$$

Sous  $H_0$ ,  $F$  suit une loi de Fisher à  $(k-1, T-k)$  ddl :  $F \sim F(k-1, T-k)$ .  
On rejette l'hypothèse nulle pour de grandes valeurs de la statistique de test.

### Significativité d'un coefficient de regression

On veut tester qu'un coefficient  $\beta_i$  vaut une certaine valeur de  $\beta_{i0}$

$$\begin{aligned} H_0 : \beta_i = \beta_{i0} \text{ contre } H_1^A : \beta_i \neq \beta_{i0} \\ \text{ou contre } H_{1B} : \beta_i > \beta_{i0} \\ \text{ou contre } H_1^c : \beta_i < \beta_{i0} \end{aligned}$$

### Statistique de test

Considérons l'expression de  $\hat{\beta}$  dans (3.9). Un sous vecteur d'un vecteur gaussien est lui-même gaussien, donc  $H_0, \hat{\beta}_i \sim N(\beta_{i0}, \sigma_0^2)$  où  $\sigma_0^2$  est l'élément  $(i, i)$  de  $(\sigma_u^2(X'X)^{-1})$ .  
Sous les quatre présupposés et si  $H_0$  est vérifiée,  $T_0 = (\hat{\beta}_i - \beta_{i0})/s(\hat{\beta}_i) \sim T(TK)$ , loi de Student à  $T-K$  ddl.

Si  $T-K > 10$ , la loi de  $T_0$  est très proche d'une loi  $N(0, 1)$  et nous nous contenterons de cette approximation par loi normale.

Sous  $H_0$

$$T_0 = (\hat{\beta}_i - \beta_{i0})/s(\hat{\beta}_i) \sim AN(0,1) \quad (3.18)$$

Si l'on choisit  $\beta_{i0} = 0$  et que  $H_0$  est rejetée, on dit que  $\beta_i$  est significatif. En toute rigueur,  $T_0$  suit sous  $H_0$  une loi de student à  $T-k$  ddl. La statistique  $\hat{\beta}_i/s(\hat{\beta}_i)$  est appelée *t-statistique*. Des statistiques semblables sont utilisées dans les modèles de séries temporelles, ou les estimateurs suivent seulement des lois approximativement normales. Nous traiterons la prédiction à la faveur de l'exemple de la section : l'étude d'une production d'électricité photovoltaïque.

### 3.2.4 Comparaison de modèles et critères d'information

En régression linéaire comme en modélisation de séries temporelles, on est amené à choisir entre différents modèles, emboîtés ou non. Il est classique d'utiliser pour cela un critère d'information.

Le principe de ce critère est le suivant : les modèles sont estimés par maximum de vraisemblance et la comparaison des valeurs de vraisemblance n'est pertinente que si les modèles ont le même nombre de paramètres.

Les critères d'information fournissent une fonction de l'opposé de la log-vraisemblance augmenté (pénalisé) d'une fonction croissante du nombre de paramètres contenus dans le modèle par des arguments théoriques qui peuvent être très divers. Cette fonction croissante peut éventuellement croître avec le nombre d'observations, étant donné plusieurs modèles et leurs estimations, et ayant choisi un critère, on retient le modèle pour lequel le critère est minimum. Nous donnons quelques détails sur deux de ces critères.

**AIC(Critère d'Information d'Akaike).**

Ce critère s'exprime par

$$AIC(k) = -2\ln(L) + 2k \quad (3.19)$$

où  $L$  est la fonction de vraisemblance évaluée en les estimations par maximum de vraisemblance (MV) des paramètres, et  $k$  est le nombre de paramètres estimés. (on rencontre quelquefois une autre expression de l'AIC :  $(-2 \ln(L) + 2k)/T$  avec  $T$  est le nombre d'observations.) dans le cas de l'ajustement par maximum de vraisemblance de (4 :3), l'AIC prend la forme :

$$AIC(k) = T\ln(\hat{\sigma}^2) + 2k \quad (3.20)$$

où  $\sigma^2$  est l'estimation MV de  $\sigma_v^2$

**SBC (Critère Bayésien de Schwartz).**

Le SBC(Schwartz' Bayesian Criterion) ou BIC (Bayésien Information Criterion) est :  $SBC(k) = -2 \ln(L) + 2k \ln(T)$  Si les erreurs sont normalement distribuées, il prend la forme  $SBC(k) = T\ln(\hat{\sigma}^2 + k\ln(T))$  On voit que ces critères (AIC et SBC) sont formés de deux termes :

- le terme  $-2 \ln(L)$  qui est d'autant plus faible que l'ajustement par maximum de vraisemblance est bon ;
- le terme  $2k$  ou  $2k \ln(T)$  qui pénalise cette faible valeur en l'augmentant par une fonction croissante du nombre de paramètres estimés.

**3.2.5 Intervalle de confiance**

Un intervalle de confiance est un intervalle dont les bornes sont aléatoires et qui contient une constante avec une probabilité qu'on se fixe. Par exemple si  $\beta_i$  est estimé par  $\hat{\beta}_i \sim N(\beta_i, \sigma_i^2)$  un IC à  $100(1-\alpha)\%$  pour  $\beta_i$  est :

$$\hat{\beta}_i \pm q(1 - \alpha/2)\sigma_i \quad (3.21)$$

où  $(1 - \alpha/2)$  désigne le quantile d'ordre  $1 - \alpha/2$  d'une v.a.  $N(0, 1)$ . si  $\sigma_i$  est inconnu, on peut approcher cet IC par  $\beta_i \pm q(1 - \alpha/2)s_i$  si  $s_i$  est l'estimation de  $\sigma_i$ . Si maintenant on s'intéresse à une combinaison linéaire des paramètres  $\beta_i \pm q(1 - \alpha/2)s_i$ , avec  $\beta$  estimé par  $\hat{\beta} \sim N(\beta, \sum_{\beta})$ , un IC à  $100(1-\alpha)\%$  pour  $c'\beta$

$$c'\hat{\beta} \pm q(1 - \alpha/2)c' \sum_{\beta_c}$$

approché par

$$c'\hat{\beta} \pm q(1 - \alpha/2)c' \sum_{\hat{\beta}_c} \quad (3.22)$$

### 3.2.6 Prédiction

Etant donné par  $X_F$  matrice  $m * (K + 1)$  de valeurs des explications, pour  $m$  observations vérifiant l'équation (3.7) et indépendant des premières observations, les  $y$  associés vérifiant :

$$Y_F = X_F\beta + U_F \quad U_F \sim N(0, \sigma_u^2 I_m) \quad (3.23)$$

Le terme " prédiction " recouvre deux situations.

#### Estimation ponctuelle ou estimation par intervalle des composants de $E(Y_F)$

Le terme  $E(Y_F) = X_F\beta$  est un vecteur de combinaison linéaire des paramètres. On estime sans biais ce vecteur certain par

$$E(\hat{Y}_F) = X_F\hat{\beta} \quad (3.24)$$

De matrice des covariances :  $\sigma^2 u(X_F(X'X)^{-1}X'_F)$  Sous l'hypothèse de normalité, on peut calculer des *IC* pour chacune des composantes de  $E(Y_F)$ . Ainsi , nous prenons  $x_f$  une ligne de  $X_F$  et  $y_f$  la valeur correspondante de  $y$  : l'*IC* à  $(1 - \alpha)\%$  pour  $E(y_f)$  est

$$x'_f \pm q(1 - \alpha/2)\sigma_u \sqrt{x_f(X'X)^{-1}x'_f} \quad (3.25)$$

Une représentation simultanée de ces *IC* est appelée bande de confiance.

#### Prédiction ponctuelle ou prédiction par intervalle des composants de $Y_F$

On prédit le vecteur aléatoire  $Y_F = X_F\beta + U_F$  par  $E(F) = x_F\hat{\beta}$  c'est-à-dire par la quantité qui a servi à estimer son espérance (ou moyenne.) Ce prédicteur est sans biais :  $Y_F = X_F\beta + U_F$  par  $E(\hat{Y}_F) = x_F\hat{\beta}$  . c'est-à-dire par la quantité qui a servi à estimer son espérance (ou moyenne.)

Ce prédicteur est sans biais :  $E(E(\hat{Y}_F) - Y_F) = 0$ .

L'erreur de prédiction est :  $X_F\hat{\beta} - Y_F = X_F(\hat{\beta} - \beta) - U_F$  ;

Sa matrice de covariance est  $\sigma^2 u(X_F(X'X)^{-1}X'_F) + I_m$ , elle est évidemment plus " grande " que la matrice des covariances de l'estimateur de la moyenne, en ce sens que  $(X_F(X'X)^{-1}X'_F) + I_m - (X_F(X'X)^{-1}X'_F) = I_m$  est défini positive.

On peut calculer des intervalles de prédiction (*IP*) pour les composants de  $Y_F$ .

La représentation simultanée de ces (*IP*) est appelée bande de prédiction. Le (*IP*) à  $(1 - \alpha)\%$  pour  $y_f$  est :

$$x'_f \pm q(1 - \alpha/2)\sigma_u \sqrt{1 + x_F(x'X)^{-1}x'_F} \quad (3.26)$$

## 3.3 Régression non linéaire

La régression non linéaire a pour but d'ajuster un modèle non linéaire pour un ensemble de valeurs afin de déterminer la courbe qui se rapproche le plus de celle des données de  $Y$  en fonction de  $x$ . Le modèle de régression linéaire s'écrit :  $Y_i = f(X_i, \Theta) + \epsilon_i$  La

loi de probabilité sur  $\epsilon_i$  est une loi normale, centrée réduite et de variance  $\sigma_i^2$  finie ; Les  $\epsilon_i$  sont indépendants entre eux ;  $Y_i$  représente l'observation  $i$  de la variable dépendante ;  $\Theta$  représente un vecteur à  $p$  composantes de paramètres généralement inconnus ; La fonction  $f$  est la fonction de régression, la plupart du temps non linéaire. Elle dépend d'une variable réelle  $x$  et de paramètres  $\Theta$ . Dans les cas où le modèle mathématique est autre non linéaire ; il faudra linéariser les données et par le fait même le modèle mathématique. Les principaux modèles que nous pourrions linéariser sont :

-le modèle logarithmique :  $Y = a_0 + a_1 \ln X$

-le modèle exponentiel :  $Y = a_0 * a_1^X$

-le modèle puissance :  $Y = a_1 * X^{a_1}$

### 3.4 Prédiction par le modèle ARMAX

Les modèles ARMAX sont des modèles de Box-Jenkins pour les cas multivariés. Comme les régressions multiples, ils permettent l'établissement de prévisions d'une variable dépendante à partir de variables indépendantes. La forme du modèle ARMAX est :

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \mu_t, \quad t = 1, 2, \dots, T$$

$$\mu_t = \frac{1 + \Theta_1 B + \dots + \Theta_1 B^q}{1 - \phi - 1B - \dots - \phi_p B^{p_z k}}$$

où  $\mu_t$  est l'erreur de modèle de régression,  $z_t$  est un bruit blanc de variance  $\sigma_z^2$  .

Le modèle ARMAX est un modèle linéaire avec erreur corrélées, son estimation est un problème de moindres carrés généralisés.

L'estimation d'un modèle ARMAX consiste à identifier le modèle de l'erreur  $t$ , elle se déroule en deux étapes.

#### 3.4.1 Identification des résidus par la méthode des moindres carrés ordinaires MCO

Les résidus sont déterminés à partir d'une régression linéaire par la MCO de  $y_t$  sur les variables explicatives, puis un modèle ARMA doit être identifié pour ces résidus.

#### Modélisation des résidus par un modèle ARIMA

Le choix du modèle résulte d'une procédure itérative qui comporte trois étapes : identification, estimation et validation du modèle.

##### Identification du modèle

Cette phase est très difficile, car le choix des paramètres  $p$ ,  $d$  et  $q$  se fait de manière priori, où éventuellement plusieurs modèles seront examinés. la fonction d'autocorrélation partielle (PAC) nous donne une idée sur l'ordre du modèle autorégressif, et celle des autocorrélations simple (AC) nous donne une idée sur l'ordre du modèle moyenne mobile. Le paramètre  $d$  représente le nombre de différenciations pour stationnariser la série.

### Estimation des paramètres

Une fois l'étape de l'identification terminée, il faut estimer les paramètres qui sont les coefficients des polynômes  $AR$  et  $MA$  ainsi que les polynômes saisonniers  $SAR$  et  $SMA$ . La méthode d'estimation la plus utilisée est celle des moindres carrés ou bien la méthode du maximum de vraisemblance. Cette dernière a donnée beaucoup de satisfaction aussi bien pour la précision des résultats obtenus que pour la stabilité et la rapidité des calculs. Quand l'étape d'estimation est achevée, l'étape suivante va nous permettre de valider le(s) modèle(s) estimé(s).

### Validation de modèle

Il en existe de très nombreux test permettant de valider le modèle retenu, nous citons :

#### -Test de Student des paramètres

Afin que le modèle soit valide, il faut que tous les coefficients soient significativement différent de zéro (leurs probabilités soient inférieures à 0,05).

Pour vérifier cela, on applique le test de Student. Si un coefficient n'est pas significativement différent de zéro, on envisage une nouvelle spécification du modèle en supprimant l'ordre  $AR$  ou  $MA$  qui n'est pas valide ou bien on calcule la statistique de student

$$t_c = \frac{|\hat{\phi}_p|}{\sqrt{\text{var}(\hat{\phi}_p)}} \quad (3.27)$$

avec  $\phi_p$  estimé au seuil  $\alpha = 5\%$ . si  $t_c > 1,96$  (la valeur tabulée de student), le paramètre est valide.

#### -Test de Ljung-Box sur les résidus

Ljung et Box ont proposé une modification qui améliore l'approximation  $Q$ . Cette statistique est donnée par la formule :

$$Q_K^K = n(n+2) \sum_{k=1}^k \left( \frac{\varphi_k^2}{n-k} \right) < X_{0.05}^2(k) \quad (3.28)$$

Si  $Q_k < X_{1-\alpha}^2(k-p-q)$  alors la série comporte comme un bruit blanc.

Si  $Q_k > X_{1-\alpha}^2(k-p-q)$  alors la série ne comporte pas comme un bruit blanc.

### 3.4.2 Estimation du modèle ARMAX

Cette étape nécessite l'identification de la variable expliquée et les variables explicatives.

Une fois l'étape de l'identification terminée, il faut estimer un modèle de régression de la variable expliquée sur les variables explicatives en spécifiant que l'erreur suit un modèle ARMA.

### 3.4.3 Validation du modèle ARMAX

La validation d'un modèle ARMAX est pareille à celle d'un modèle ARMA. Elle concerne les tests sur les paramètres et les tests sur les résidus.

### 3.4.4 Prévision du modèle ARMAX

Une fois terminées les étapes précédentes, on arrive à l'étape de prévision.