



Journées de Recherche des ISTs et de leurs partenaires internationaux - 2 et 3 décembre 2020, Antsiranana
 « **L'innovation et le développement durable : perspectives, enjeux et défis sociétaux** »

Article 30 : Deux approches efficaces pour estimer un modèle de série temporelle

T. Rabenantenaina¹, P. Bemarisika¹, A. Totohasina¹

¹Ecole Normale Supérieure pour l'Enseignement Technique, Université d'Antsiranana

Correspondant : rabenantheophile@gmail.com

Résumé

La méthodologie de Box-Jenkins vise à formuler un modèle permettant de représenter une chronique avec comme finalité de prévoir des valeurs futures. De ce fait, l'objet de cette méthode est de modéliser une série temporelle en fonction de ses valeurs passées en utilisant des modèles ARMA, ARIMA et SARIMA. Toutefois, cette méthodologie présente une limite notable sur son étape pour estimer un modèle temporel. De plus, cette approche est utilisée pour sélectionner les modèles linéaires de Box-Jenkins seulement. Tout au long de cet article, nous essayons de résoudre ces problèmes en utilisant la propriété qui relie les modèles ARMA et GARCH et la spécification automatique des modèles ARMA (*i.e.* des critères d'information). Dans ce cas, nous élaborons une nouvelle approche qui améliore cette méthodologie. Cette nouvelle approche améliore la phase d'identification de ladite méthodologie (*i.e.* approche) de Box-Jenkins. Ensuite, nous proposons aussi une autre approche qui dépend de l'approche de Box-Jenkins pour estimer les modèles non linéaires comme : ARCH et GARCH. On termine ce travail par quelques applications sur de données réelles.

Mots clés : Méthodologie de Box-Jenkins, modèle de série temporelle, critère d'information, autocorréogrammes, stratégie d'estimation

1. INTRODUCTION

L'analyse des séries temporelles unidimensionnelles par la méthode de Box-Jenkins comporte traditionnellement cinq étapes : transformation, identification de la forme de modèle la mieux adaptée à la série étudiée, estimation des paramètres, validation du modèle retenu et prévision. Sous cette forme, la méthode est déjà d'un emploi très général et son intérêt n'est plus à démontrer. Cependant, la phase d'identification révèle, en pratique, d'être une étape délicate. Elle repose sur l'analyse des autocorrélographes de la série et il arrive malheureusement assez souvent qu'ils soient relativement peu typiques. De plus, cette méthodologie ne sélectionne pas les modèles non linéaires. Cette non linéarité peut concerner : les modèles TAR "Threshold-AR", les modèles EXAR (AR Exponentiel), les modèles de type ARCH [10], etc. Dans ce papier, nous proposons une nouvelle approche qui améliore l'étape d'identification. Elle se repose sur la spécification automatique des modèles ARMA, qui est aussi appelée *critère de comparaison des modèles*. Nous proposons aussi une autre approche qui sélectionne les modèles non linéaires comme ARCH et GARCH et dépendant de l'algorithme de Box-Jenkins. Le reste de ce papier est organisé comme suit. La section 2 expose les matériels et méthodes, résultats et discussions. Une conclusion est donnée dans la section 3.

2. CONTENU

2.1. Matériels et méthodes

2.1.1. Méthodologie de Box-Jenkins

Box et Jenkins [4], [5] ont promu une méthodologie consistant à modéliser les séries temporelles univariées au moyen des processus ARMA. Ces processus sont parcimonieux et constituent une bonne approximation de processus plus généraux pourvu que l'on se restreigne au cadre linéaire. Les modèles ARMA donnent souvent de bons résultats en prévision et ont bénéficié de la vague de scepticisme quant à l'intérêt des gros modèles économétriques. Cette méthodologie comporte essentiellement les étapes suivantes : transformation des données afin de stabiliser la variance (\log , $\sqrt{\quad}$, ...) et différenciation des données pour les stationnariser, visualisation des ACF (fonction d'autocorrélation) et des PACF (fonction d'autocorrélation partielle) empiriques pour identifier les paramètres p et q appropriés, estimation des paramètres du (des) modèle(s) sélectionné(s) en utilisant la méthode de maximum de vraisemblance, etc, diagnostic et tests d'adéquation du modèle et prévision : la dernière étape consiste à la prévision des valeurs futures à travers le modèle retenu. La première étape a pour objet de transformer la série de données afin d'éliminer d'éventuelles causes de non stationnarité telles qu'une tendance et/ou une saisonnalité. Pour cela, une transformation peut être réalisée sur la série, ainsi qu'une différenciation simple et/ou saisonnière. Les tests tels que test de KPSS,

test de Phillips et Perron, test de Dickey et Fuller, etc, permettent de conclure quant à la stationnarité ou non de la série étudiée. Il s'agit ensuite d'identifier pour les données transformées et rendues stationnaires, un modèle de type ARMA. L'examen des autocorrélations empiriques et des autocorrélations empiriques partielles sur les graphiques ACF et PACF permet la plupart du temps de sélectionner des modèles à tester. Il convient de noter qu'à ce stade, plusieurs modèles peuvent parfois être envisagés. Les paramètres du modèle doivent ensuite être estimés avant de tester l'adéquation du modèle aux données pour pouvoir envisager de faire de la prévision. Voici le résumé de cette méthodologie sous forme d'algorithme.

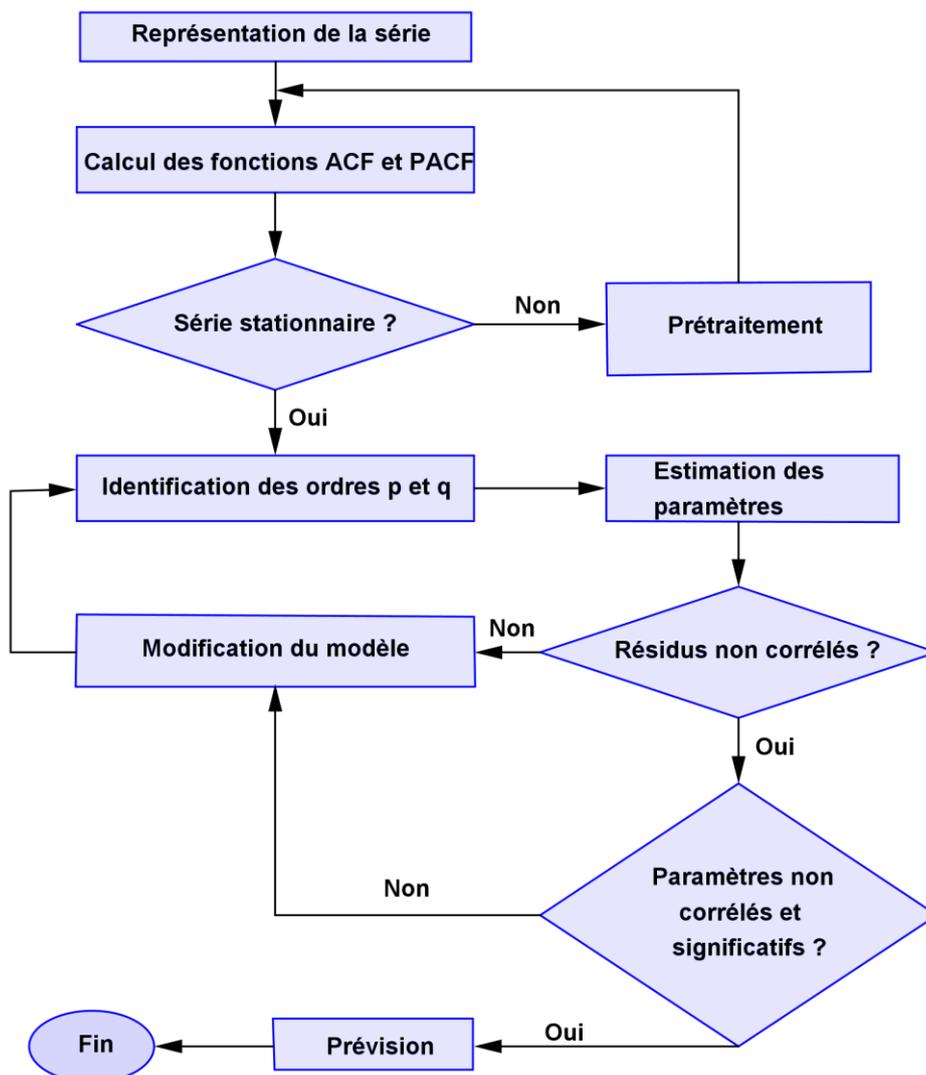


Figure 1 : Algorithme de la méthodologie de Box-Jenkins.

2.1.2. Spécification automatique des modèles ARMA

Une autre approche de la sélection du modèle consiste à utiliser la spécification des modèles ARMA. Cette spécification est appelée aussi **critère de comparaison des**

modèles. L'identification des processus ARMA (p,q) à partir de la visualisation des autocorrélations n'est pas chose facile. Pour résoudre ce problème, certains auteurs ([1], [2], [9], ...) ont développé des critères, issus des concepts de la théorie de l'information. Ces critères sont basés sur la maximisation de la fonction log –vraisemblance afin d'obtenir un critère d'entropie.

Critère d'information d'Akaike : Maintenant, nous allons voir l'un de ces critères, c'est-à-dire le critère d'AIC (Akaike's Information Criterion).

Définition 2.1. Le critère le plus connu, l'« AIC » est défini pour un modèle à k paramètres, par la relation suivante [1], [2], [6] :

$$AIC(k) = -2 \log(\text{vraisemblance maximisée}) + 2k \quad (1)$$

ou

$$AIC(k) = -2(\text{maximum de } \log \text{ vraisemblance}) + 2k \quad (2)$$

Il faut donc déterminer la log –vraisemblance du modèle ainsi que son maximum. Or, on sait que, dans le cas des moindres carrés :

$$(y = X\beta + \varepsilon) \quad (3)$$

la log –vraisemblance approchée est :

$$L(\beta) = Cte - \frac{1}{2\sigma_\varepsilon^2} S(\beta) \quad (4)$$

où $S(\beta)$ est la somme des carrés des résidus. Pour un processus ARMA (p,q) , β sera un ensemble de $k=p+q+1$ paramètres :

$$\beta = (\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) \text{ et } S(\beta) = \sum_{t=p+1}^T \varepsilon_t^2 \text{ où } \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad (5)$$

Si on suppose que $\hat{\beta}$ est l'estimateur du maximum de vraisemblance de β , on peut écrire :

$$L(\beta) = L[\hat{\beta} + (\beta - \hat{\beta})], \text{ sachant que } \hat{\beta} = (\hat{\mu}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_p) \quad (6)$$

On montre que la log –vraisemblance estimée pour T observations pouvait s'écrire :

$$L(\hat{\beta}) = -\frac{T-p}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} S(\hat{\beta}) \quad (7)$$

où σ_ε^2 est estimé par :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T-p} S(\hat{\beta}) \quad (8)$$

Commentaire : $\hat{\sigma}_\varepsilon^2$ est l'estimateur du maximum de vraisemblance conditionnelle de la variance résiduelle, donc le maximum de $L(\hat{\beta})$ est donné par :

$$L(\hat{\beta}) = -\frac{T-p}{2} \log(\hat{\sigma}_\varepsilon^2) - \frac{T-p}{2} \quad (9)$$

On ignore le terme constant puisqu'il est indépendant de β . Le critère AIC s'écrit alors :

$$AIC(k) = (T - p) \log(\hat{\sigma}_\varepsilon^2) + 2k \quad (10)$$

Si l'on trace sur un graphe AIC(k) en fonction des différentes valeurs de k , on obtient en général un maximum pour une certaine valeur de k . Ce minimum, appelé MAIC

(minimum AIC) donne donc le nombre de paramètres adéquats pour estimer le modèle ARMA(p,q). La figure 2 ci-dessous montre ce phénomène.

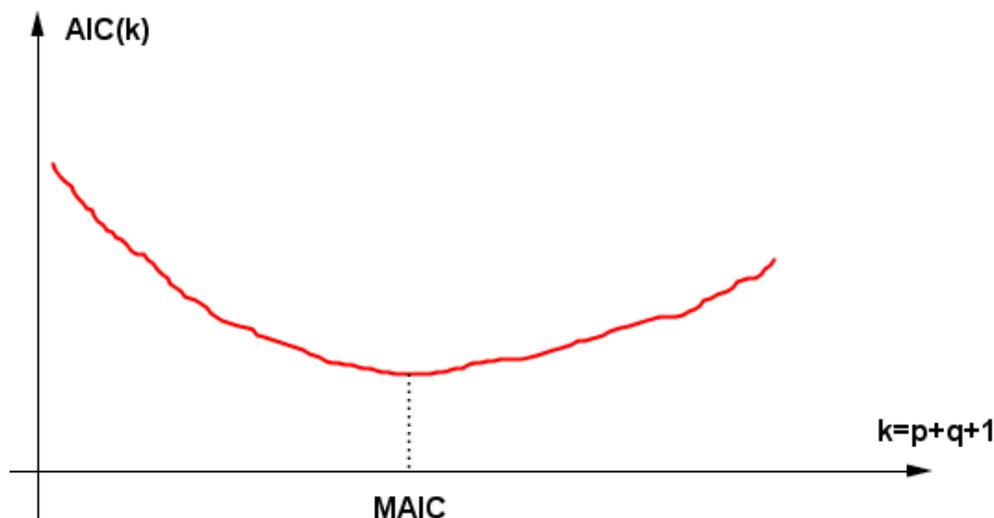


Figure 2 : Représentation du MAIC

Ce critère établit un compromis entre l'adéquation du modèle sur les n valeurs x_t observées dans le passé qui ont servi à estimer le modèle et sa complexité calculée en fonction du nombre de paramètres. Il est en cela conforme au principe de parcimonie veillant à construire des modèles les moins complexes possibles afin d'éviter la redondance. Ainsi, le modèle ayant la plus petite valeur du critère devra être choisi.

2.1.3. Relation entre les modèles non linéaires : ARCH, GARCH et ARMA

2.1.3.1. Modèles autorégressifs conditionnellement hétéroscédastiques : processus ARCH

Le modèle ARCH développé par [7], [13] est tel que la variance conditionnelle dépend du niveau des valeurs passées. Dans la théorie des processus ARMA, la variance d'une série est, entre autres, déterminée par la variance du processus des innovations.

Définition 2.2. Un processus ARCH(p) est défini par :

$$X_t = \varepsilon_t \tag{11}$$

$$\varepsilon_t | X_{t-1}, X_{t-2}, \dots \sim \mathcal{N}(0, \sigma_t^2) \tag{12}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 \tag{13}$$

La variance conditionnelle dépend du temps : si les valeurs précédentes sont grandes (en valeur absolue), la variance sera grande, et inversement. Ainsi, si on observe un choc dans la série (valeur anormalement grande), elle sera suivie d'une période de haute volatilité, dont la durée dépend de l'ordre p du modèle ARCH.

2.1.3.2. Processus GARCH et propriétés

[3], [13] ont proposé une généralisation du modèle ARCH qui s'est montré très utilisé en pratique. Il a proposé de remplacer la fonction de variance conditionnelle (11), (12), (13) par σ_t^2 définie par :

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2 \quad (14)$$

où les paramètres α_i et β_j sont tels que

$$\left\{ \begin{array}{l} \alpha_0 > 0; \\ \dots \\ \alpha_i \geq 0, \quad \text{pour } i = 1, \dots, p; \\ \dots \\ \beta_j \geq 0, \quad \text{pour } j = 1, \dots, q. \end{array} \right.$$

Notons que (14) est en fait une généralisation de (11), (12), (13), et le modèle (14) est appelé GARCH(p, q). Pour motiver l'introduction des processus GARCH, on peut réécrire (14) à l'aide des opérateurs $A(\cdot)$ et $\Theta(\cdot)$. Dans ce nouveau contexte, ces opérateurs sont définis par

$$A(z) = \alpha_1 z + \dots + \alpha_p z^p \quad (15)$$

et

$$\Theta(z) = \beta_1 z + \dots + \beta_q z^q \quad (16)$$

On peut donc écrire :

$$\sigma_t^2 = \alpha_0 + A(B)X_t^2 + \Theta(B)\sigma_t^2 \quad (17)$$

où B est l'opérateur de retard. En utilisant le théorème de causalité et d'inversibilité des modèles ARMA(p, q), on voit donc directement que si les racines du polynôme $1 - \Theta(z)$, $z \in \mathbb{C}$, sont en dehors du cercle unité du plan complexe $\{z : |z| \leq 1\}$, alors on peut réécrire (14) comme suit :

$$\sigma_t^2 = b_0 + \sum_{i=1}^{\infty} b_i X_t^2 \quad (18)$$

Cette dernière relation montre qu'un processus GARCH(p, q) est un processus ARCH d'ordre infini. Ainsi, la généralisation des processus ARCH aux processus GARCH est similaire à la généralisation des processus autorégressifs aux processus ARMA. Par cette comparaison, on voit donc que les processus GARCH peuvent formellement représenter de façon plus parcimonieuse des processus ARCH contenant un nombre élevé de paramètres. Bien entendu, l'absence de corrélations n'implique pas que les valeurs X_t d'un processus GARCH soient indépendantes. En effet, en notant $\eta_t = X_t^2 - \sigma_t^2$, le carré des valeurs du processus s'écrivent ainsi :

$$X_t^2 = \eta_t + \sigma_t^2 \quad (19)$$

$$X_t^2 = \alpha_0 + \sum_{i=1}^p X_t^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \eta_t \quad (20)$$

Si on note $m = \max(p, q)$ et si on convient que $\alpha_i = 0$, pour $i > p$ et $\beta_i = 0$, pour $i > q$, alors cette dernière expression peut s'écrire sous la forme suivante :

$$X_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i - \beta_i) X_t^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \eta_t \quad (21)$$

$$X_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i) X_t^2 + \sum_{j=1}^q \beta_j (\sigma_{t-j}^2 - X_t^2) + \eta_t \quad (22)$$

et finalement, par la définition du processus η_t :

$$X_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i) X_t^2 - \sum_{j=1}^q \beta_j \eta_{t-j} + \eta_t \quad (23)$$

Puisque $E(\eta_t | \mathcal{F}_{t-1}) = 0$, la loi des espérances itérées implique que le processus η_t est de moyenne nulle et est non corrélé. En conséquence, (23) indique que $\{X_t^2\}$ a une représentation ARMA(m, q), et $\{\eta_t\}$ joue le rôle du processus des innovations dans cette représentation. Cette observation amène deux conditions immédiates précisées ci-dessous.

Propriété 2.1. 1. Bien que les valeurs X_t d'un processus GARCH soient non corrélées, il existe des dépendances non linéaires entre les observations, puisque le carré des observations se comporte formellement comme un processus ARMA.

2. Pour identifier les paramètres p et q d'un processus $X_t \sim \text{GARCH}(p, q)$, on peut utiliser les fonctions d'autocorrélation et d'autocorrélation partielle du processus $\{X_t^2\}$ suivant la même procédure utilisée pour trouver le nombre de paramètres d'un processus ARMA. Plus précisément, on trouve d'abord la représentation ARMA(m, q) pour $\{X_t^2\}$ où, rappelons-le, $m = \max(p, q)$. Au cas où $m = q$, il faut, bien évidemment, passer par une étape supplémentaire pour en pouvoir déduire l'ordre $p \leq q$ du processus GARCH(p, q) qui suggère la possibilité d'effectuer des tests d'hypothèse successifs de significativité des paramètres $\alpha_q, \dots, \alpha_1$ du processus ARMA(m, q) (sont-ils significativement non nuls ?).

Remarque 2.1. Un GARCH($p, 0$) est un ARCH(p).

2.2. Résultats

2.2.1. Une nouvelle approche d'estimation d'un modèle temporel

Nous savons qu'à partir de critère d'information, nous pouvons identifier le modèle ARMA le plus fiable pour une base stationnaire quelconque. Dans ce cas, nous élaborons une technique d'estimation des ordres p et q appelée *AicSelection*, fonction créée sous **R** et dépendant de critère d'information AIC. Le pseudo-code de cette fonction est donné ci-dessous.

```

AicSelection=function(time.serie,max.iter=Inf, trace=F, order=c(0,0))
{if (!is.ts(time.serie)) { if (!is.numeric(time.serie)) {
  stop("'time.serie' doit être un objet de type 'ts' ")}}
else {time.serie=ts(time.serie)}}
if (max.iter != Inf && round(max.iter) !=max.iter) {
  stop("'max.iter' doit être entier ou égal à 'Inf' ")}}
if (trace !=F && trace !=T) {stop("'trace' doit être un booléen")}}
if (!is.numeric(order) || round(order) !=order || length(order) !=2){
  stop("'order' doit être un vecteur de 2 entiers positifs")}}
p.opt=order[1]
q.opt=order[2]
black.list=matrix(c(p.opt, q.opt), nrow=2)
fit.opt=arima(time.serie, order=c(p.opt, 0, q.opt))
if (trace=="F"){print(paste(model, "(", p.opt, ",", q.opt, "): ", fit.opt$aic, sep=" "))}}
n.iter=0
while (n.iter < max.iter) {n.iter=n.iter+1
  fit.best=fit.opt
  p.best=p.opt
  q.best=q.opt
  for(i in (p.opt-1):(p.opt+1)) {
    for (j in (q.opt-1):(q.opt+1)){
      if (!any(apply(black.list==c(i,j), 2, all)) && min(i,j)>=0) {
        black.list=cbind(black.list, c(i,j))
        fit.temp=arima(time.serie, order=c(i, 0, j))
        if (trace=="F"){print(paste(model, "(", i, ",", j, "): ", fit.temp$aic, sep=" "))}}
        if (fit.temp$aic < fit.best$aic) {fit.best=fit.temp
          p.best=i
          q.best=j}}}}
  if (fit.best$aic < fit.opt$aic) {
    p.opt=p.best
    q.opt=q.best
    fit.opt=fit.best}
  else{break}}
ans=list(order=c(p.opt, q.opt), aic=fit.opt$aic)
ans}

```

Et à partir de cette technique, nous élaborons une méthodologie qui améliore celle de Box-Jenkins et qui est basée sur les critères d'informations. Dans ce cas, nous utilisons la technique *AicSelection* pour identifier le couple (p,q) le plus adéquat pour remplacer l'étape d'identification proprement dit. Le pseudo-code est présenté dans l'algorithme 1 ci-dessous.

Algorithme 1 : Algorithme d'estimation des modèles linéaires

Entrée : Base de données

Sortie : Valeurs futures

- 1: $VF = \emptyset$ (Initialisation)
 - 2: Faire la représentation graphique de données
 - 3: Calculer la fonction ACF (à présenter graphiquement aussi)
 - 4: **Tant que** La série n'est pas stationnaire **faire**
 - 5: Prétraitement de données
 - 6: Calculer la fonction ACF
 - 7: **Fin Tant que**
 - 8: Identifier les ordres p et q en utilisant la fonction « *AicSelection* »
 - 9: Estimer les paramètres du modèle retenu
 - 10: Faire des prévisions et mettre dans VF ces valeurs
-

2.2.2. Utilisation de la nouvelle approche pour estimer les modèles non linéaires ARCH et GARCH

Dans l'algorithme de Box-Jenkins, les modèles sélectionnés sont des modèles linéaires. Mais plus récemment, certains auteurs ([8], [10], [11], [12] et [14]) ont suivi des approches à savoir : paramétriques, non paramétriques, par maximum de vraisemblance et primes d'asymétrie pour identifier des processus non linéaires. Mais à partir de la propriété ci-dessus (propriété 2), on peut identifier des modèles GARCH(p,q) et ARCH(p) ou GARCH($p,0$) à l'aide des modèles ARMA, et en voici les étapes à suivre :

- Si on a une base de données ($data=time.serie$), on doit avoir les carrées des composantes de $data$ ($dax = data^2$).
- On cherche à ajuster un modèle ARMA(m,q) avec $m \geq q$ sur dax et on choisit (m,q) de manière à minimiser la quantité AIC ou BIC.
- On propose après un modèle GARCH(p,q) qui modélise $data$. On montre que les deux premières étapes de cette technique d'estimation sont l'approche de Box-Jenkins. Par contre, nous avons déjà amélioré cette dernière. Alors, pour faciliter cette approche d'estimation, il est mieux de suivre la nouvelle approche.

Ainsi, nous allons élaborer un algorithme pour estimer les modèles non linéaires et le pseudo-code est présenté dans l'algorithme 2 ci-dessous.

Algorithme 2 : Algorithme d'estimation des modèles non linéaires

Entrée : Base de données

Sortie : Valeurs futures

1: $VF = \emptyset$ (Initialisation)

2: Elever au carrée toutes les composantes de données

3: Faire la représentation graphique d'une base élevée au carrée

4: Identifier les ordres m et q de $X_t^2 \sim \text{ARMA}(m,q)$ en utilisant la fonction «AicSelection»

avec $m = \max(p, q)$

5: **Si** $m > q$ **alors**

6: $p = m$

7: **Sinon**

8: Identifier l'ordre p tel que $X_t \sim \text{GARCH}(p, q)$ avec $p = \overline{1, q}$

9: **Fin Si**

10: Estimer les paramètres de $X_t \sim \text{GARCH}(p, q)$

11: Faire les prévisions et mettre dans VF ces valeurs

12: **Retourner** Valeurs futures VF

2.4. Discussions

Dans cette section, nous allons prendre un exemple afin d'illustrer le fonctionnement de la nouvelle approche dont nous avons parlé ci-dessus et de pouvoir faire la comparaison à celle de la méthodologie de Box-Jenkins. Pour ce faire, on prend deux bases de données et on va les modéliser.

2.4.1. Applications expérimentales

(i) Premier exemple

Nous allons considérer la base de données nommée « *receipt⁴* », il s'agit des revenus nets du gouvernement fédéral des U.S.A. pendant dix ans, à partir de janvier 1976 (données mensuelles, de janvier 1976 à décembre 1986). Supposons observée la série $(X_t, t \in \mathbb{Z})$ que nous souhaitons modéliser à partir de ces données. Ignorons les étapes qui se ressemblent entre l'approche de Box-Jenkins et la nouvelle approche. Nous faisons seulement les étapes différentes entre eux. Après avoir effectué quelques transformations, on obtient la série différenciée $(Y_t = Z_t - Z_{t-1})$, avec $(Z_t = \log(X_t))$ qui est stationnaire. Et sa présentation graphique est donnée par la figure 3 ci-dessous :

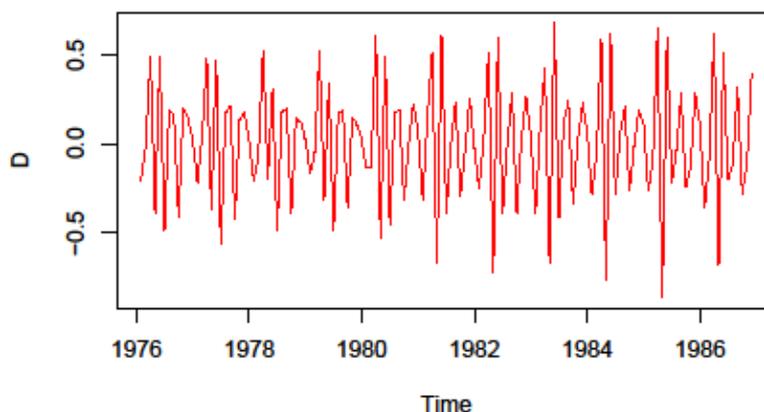


Figure 3 : Représentation graphique de la série stationnaire

Identification des ordres p et q

a) Pour la nouvelle approche : On utilise la technique/ la fonction *AicSelection* pour identifier le modèle le plus significatif. Pour notre cas, voici le résultat obtenu après avoir utilisé cette fonction :

```
$order
[1] 2 1
$aic
[1] -66.63594
```

⁴ Source : Université de Toulouse I

Donc, le modèle $ARMA(2,1)$ s'avère ainsi plus significatif selon le critère d'AIC pour modéliser la série $(Y_t)_{t \in \mathbb{Z}}$.

b) Pour la méthodologie de Box-Jenkins : On doit utiliser les graphes des autocorrélogrammes afin de pouvoir identifier les listes des ordres p et q candidats. Et la présentation graphique des fonctions ACF et PACF est donnée par la figure 4.

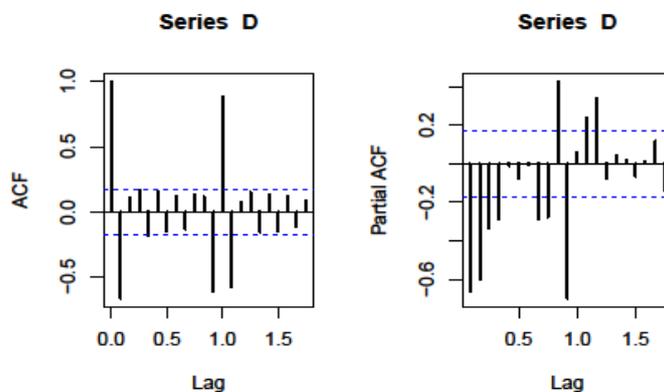


Figure 4 : Représentation graphique des fonctions ACF et PACF de la série (Y_t)

Compte tenu de l'allure des autocorrélogrammes de (Y_t) , on peut lister les ordres possibles de p et q : $p=1, 2, 3, 4$ et $q=0,1$. Donc les listes de couples (p,q) possibles sont : $(1,0), (2,0), (3,0), (4,0), (1,1), (2,1), (3,1)$ et $(4,1)$. Dans ce cas, on doit estimer les paramètres et faire le test de normalité des résidus et le test de significativité des coefficients de chaque couple possible afin d'avoir le couple plus significatif.

(ii) Deuxième exemple

Maintenant, nous allons prendre une autre base de données nommée « RBS^5 », il s'agit de niveau action de la Royal Bank of Scotland (quotidien à partir du 03 janvier 2000 au 11 janvier 2009) afin d'appliquer la deuxième approche. En faisant les transformations suivantes : $dax <- -\log(rbs)$ et $Dat = \text{diff}(dax, \text{lag}=1, \text{difference}=1)$, on obtient une nouvelle base nommée « Dat ». Notre objectif est de modéliser cette nouvelle base par les modèles non linéaires comme ARCH ou GARCH. Soit $F = Dat^2$ (les carrés des composantes de Dat). Supposons que (D_t) la série à modéliser. Maintenant on cherche à ajuster un modèle $ARMA(m,q)$ sur la série (D_t^2) avec $m = \max(p,q)$. Et sa présentation graphique est donnée par la figure 5.

⁵ Source : Université de Toulouse I

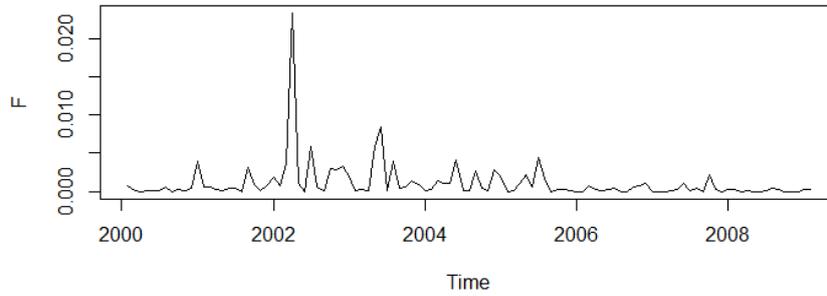


Figure 5 : Représentation graphique de la série (D_t^2) .

Identification des ordres m et q : comme nous l'avons signalé dans la section 2.2.2, on peut suivre la méthodologie de Box-Jenkins pour faire les deux premières étapes de l'approche pour estimer les modèles non linéaires.

a) Pour la nouvelle approche : On utilise la technique/ la fonction *AicSelection* pour identifier le modèle le plus significatif. Pour notre cas, voici le résultat obtenu après avoir utilisé cette fonction :

```
$order
[1] 1 1
$aic
[1] -987.8926
```

Donc, le modèle $ARMA(1,1)$ est plus significatif selon le critère d'AIC pour modéliser la série $(D_t^2)_{t \in \mathbb{Z}}$. Comme ici $m=q=1$, alors $p=1$. Ainsi, le modèle $GARCH(1,1)$ s'avère le plus significatif pour modéliser la série $(D_t)_{t \in \mathbb{Z}}$.

b) Pour la méthodologie de Box-Jenkins : On doit utiliser les graphes des autocorrélogrammes afin de pouvoir identifier les listes des ordres m et q candidats. Et la présentation graphique des fonctions ACF et PACF est donnée par la figure 6.

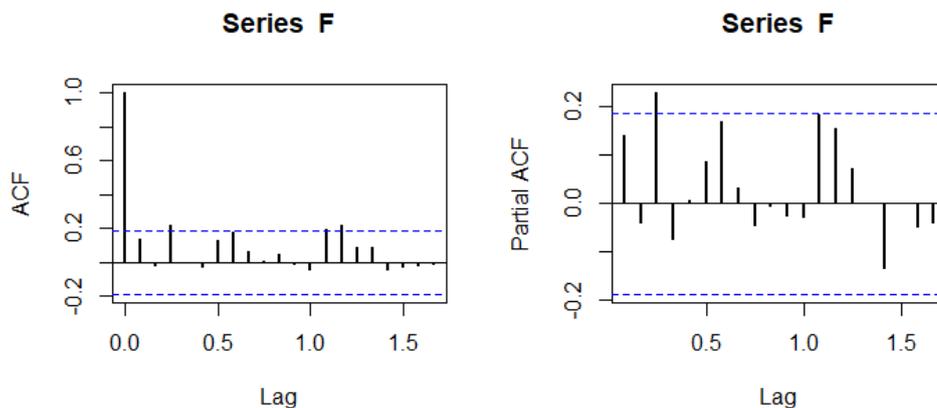


Figure 6 : Représentation graphique des fonctions ACF et PACF de la série (D_t^2) .

Compte tenu de l'allure des autocorrélogrammes, nous pouvons lister les ordres m et q de la série (D_t^2) , tels que $m=1, 2, 3$ et $q=0, 1, 2, 3$. Et les couples possibles de (m,q) avec $m = \max(p, q)$ sont : $(1,0)$, $(1,1)$, $(2,0)$, $(2,1)$, $(2,2)$, $(3,0)$, $(3,1)$, $(3,2)$ et $(3,3)$. Il reste à estimer les paramètres et faire le test de normalité des résidus et le test de significativité des coefficients de chaque couple possible afin d'avoir le couple plus significatif. Et après avoir obtenu le couple significatif pour le modèle ARMA, on propose le modèle GARCH qui convient à la base de données (Dat).

2.4.2. Comparaison

D'après l'application expérimentale que nous avons faite précédemment, on voit bien qu'il y a une différence entre la nouvelle approche et la méthodologie de Box-Jenkins. Parmi les étapes de la méthodologie de Box-Jenkins, l'étape d'identification est plus difficile et délicate, car elle se repose sur l'analyse des fonctions d'autocorrélation et d'autocorrélation partielle. Mais parfois, il est difficile d'identifier l'ordre p et q à partir des autocorrélogrammes. De plus, le choix du couple le plus fiable parmi les candidats demande plusieurs tests à faire. Donc, ces étapes ont un coût exponentiel. Tandis que dans la nouvelle approche, on utilise la fonction appelée *AicSelection* sous **R**, pour estimer le couple (p,q) le plus fiable pour une base stationnaire quelconque. En fait, il y a une diminution du coût par rapport à celle de Box-Jenkins. Pour cela, les étapes à suivre sont : validation des données, préparation des données : stationnarisation, identification, estimation des paramètres et prévisions.

3. CONCLUSION

Dans ce papier, nous avons proposé deux approches d'estimation d'un modèle de série temporelle. La première, c'est l'approche d'estimation d'un modèle linéaire, qui utilise la spécification automatique des modèles ARMA dans le but d'améliorer l'approche de Box-Jenkins. Et la deuxième, c'est l'approche d'estimation d'un modèle non linéaire, qui dépend de la première approche. Cette dépendance est une conséquence de la relation entre ARMA et GARCH. Par rapport aux approches existantes dans la littérature, ces deux approches sont faciles à utiliser dans la pratique, car il y a une diminution du coût en termes d'estimation d'un modèle de série temporelle. Cela apporte aussi des avantages sur l'approximation réaliste des valeurs futures de séries temporelles dans plusieurs domaines comme : finance, économie, météorologie, etc.

Références

[1] H. Akaike, *A new look at the statistical model identification*, *IEEE Transactions on Automatic Control*, vol.19, n°6, pp.716-723, 1974

- [2] H. Akaike, Information theory and an extension of the maximum likelihood principle, In Second International Symposium in Information Theory, pp.276-281, 1973
- [3] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, In Journal of Econometrics, Vol.31, pp.307-327, 1986
- [4] G. E. P. Box et G. M Jenkins, Time series analysis : Forecasting and control, San Francisco, Calif. : Holden-Day, 1970
- [5] P. J. Brockwell et R. A. Davis, Introduction to time series and forecasting, Springer Texts in Statistics, Springer-Verlag, New York, second édition, 2002
- [6] K. P. Burnham et D. R. Anderson, *Model Selection and Multimodel Inference : A Practical Information-Theoretic Approach*, Springer-Verlag, 2002
- [7] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, In Econometrica, Vol.50, pp.987-1007, 1982
- [8] F. Jianqing et Y. Qiwei, Nonlinear Time Series : Nonparametric and Parametric Methods, Springer series in statistics, 2003
- [9] G. Schwarz, Estimating the dimension of a model, Ann. Statist, Vol.6, pp.461-464, 1978
- [10] M. P. Simon, Nonlinear time series modelling : An introduction, Federal Reserve Bank of New York, 33 Liberty St. New York, 1999
- [11] S. Sjöholm, Heteroscedasticity Models and their Forecasting Performance, Department of Mathematics : Uppsala University, septembre 2015
- [12] H. Tong, Non-linear time series : A dynamical systems approach, Oxford University Press, 1990
- [13] R. S. Tsay, Analysis of Financial Time Series, John Wiley & Sons, Inc. ISBN : 0-471-41544-8, 2002
- [14] A. A. Yacouba, Estimation d'un modèle ARCH-GARCH avec primes d'asymétrie, Master's thesis, Université Laval, Québec, Canada, 2013