Etude du pouvoir estimatif de la méthode Geweke Porter-Hudak sur les modèles ARFIMA : Application sur la température de l'aire de la ville d'Alger

Mr. BENYAMMI Youcef[1]

Résumé :

Ce papier présente une méthode de prévision d'une série chronologique basé sur un modèle à mémoire longue à savoir le processus autorégressif- moyenne mobile fractionnairementintégré (noté ARFIMA par la suite). Le paramètre d'intégration fractionnaire dans le modèle ARFIMA est estimé par une méthode semi-paramétrique de Geweke Porter-Hudak (GPH par la suite). Nous essayons de déterminer les différents facteurs influant sur l'estimation par GPH du paramètre d'intégration fractionnaire (d) des séries simulées par la méthode de Monte Carlo, puis on utilise cette méthode d'estimation pour prédire la série temporelle de température de l'aire de la ville d'Alger.

Mots clés : mémoire longue, ARFIMA, GPH, prévision, température de l'aire.

Introduction :

Une grande partie d'analyse des séries chronologiques considère le cas où l'ordre d'intégration, d, est un entier. Si une série est intégrée d'ordre un ou plus, cette série n'est pas stationnaire, et sa fonction d'autocorrélation (ACF) diminue linéairement. Et si elle est intégrée d'ordre zéro, son ACF montre une décroissance exponentielle.Donc, on peutdire que des observations séparées par une longue période sont indépendantes. Beaucoup de travaux ont discuté l'analyse d'un tel comportement dans des détails considérables. Néanmoins, beaucoup

[1] Maitre-assistant à l'université d'Alger 03

de séries chronologiques empiriquement observées, semble être stationnaire (même après certaine différenciation), montre une dépendance entre les observations éloignées qui(bien que petite) est nullement négligeable.Ces séries se caractérisent par une fonction d'autocorrélation qui décroît hyperboliquement. Ce type de comportement est appelé mémoire longue. Ce phénomène est apparu dans les années 1895 à partir des observations de l'astronome New Comb puis du chimiste Student (1927). Le domaine qui a été très certainement à l'origine de l'essor des modèles à mémoire longue est l'hydrologie, avec les travaux fondateurs de Hurst (1951) sur les crues du Nil. Ses travaux ont montré que certaines séries présente une structure de corrélation particulière (mémoire longue), et il a introduit un outil statistique qui permet de détecter la mémoire longue dans une série chronologique et qui porte son nom : exposant de Hurst noté H.

Ces travaux ont été suivis et approfondi par d'autres chercheurs qui ont élaboré d'autres modèles qui caractérisent ce phénomène, comme les processus auto-similaire (self-similar) ainsi que des processus de séries temporelles ont été développés afin de rendre compte des propriétés atypiques de la fonction d'autocorrélation et de la densité spectrale. Ces processus, appelés **ARFIMA** (Auto-regressiveFractionallyItegratedMovingAverage) ; ont été introduits pour la première fois par Granger et Joyeux (1980). Ces processus constituent un prolongement des modèles ARIMA, dans lesquels le coefficient d'intégration prend des valeurs réelles ($d \in R$), on l'appelle le coefficient d'intégration fractionnaire. Les propriétés statistiques de ces processus ont fait l'objet de plusieurs recherches (Granger et Joyeux (1980) et Hosking (1981)) et elles sont maintenant bien connues.

Dans notre travail on commence de présenter une définition de la notion de mémoire longue ainsi que la définition, les propriétés et la méthode d'estimation de Geweke Porter-Hudak (GPH par la suite) du processus ARFIMA. Puis on effectuera une application théorique de la

méthode d'estimation GPH basé sur des simulations de Monté Carlo, puis on consacre la dernière partie à l'application de la méthode d'estimation proposée sur la série journalière de température de l'aire dans la ville d'Alger de l'année 2012.

1. Les processus à mémoire longue :

1.1 Définition des processus à mémoire longue :

Les processus à mémoire longue peuvent être définis de façon équivalente dans le domaine spectral et le domaine temporel.

1.1.1 Dans le domaine fréquentiel :

Les processus à mémoire longue sont caractérisés par une densité spectrale s'accroissant sans limite quand la fréquence tend vers zéro.

Définition :

Un processus stationnaire $\{X_t\}_{t\in Z}$ est un Processus à mémoire longue s'il existe un nombre réel $\beta, 0 < \beta < 1$ et une constante c', c' > 0 vérifiant :

$$\lim_{\lambda\to 0} \frac{f(\lambda)}{c'|\lambda|^{-\beta}} = 1 \qquad (1.1)$$

$f(\lambda)$: la densité spectrale du processus $\{X_t\}_{t\in Z}$ à la fréquence λ

On en déduit immédiatement que$f(\lambda) \sim c' \left| \lambda \right/^{-\beta}$, quand$\lambda \to 0$. Ainsi la densité spectrale exhibe un pôle à la fréquence zéro, contrairement à la densité spectrale des processus à mémoire courte qui est finie et positive aux basses fréquences.

1.1.2 Dans le domaine temporel

Les processus à mémoire longue sont caractérisés par une fonction d'autocorrélation décroissante hyperboliquement au fur et à mesure que le retard s'accroît, à l'encontre des processus à mémoire courte où elle décroît exponentiellement.

Définition :

Un processus stationnaire $\{X_t\}_{t\in Z}$ est un Processus à mémoire longue s'il existe un nombre réel $\alpha, 0 < \alpha < 1$et une constante c, c > 0 vérifiant :

$$\lim_{k\to\infty} \frac{\rho(k)}{c\, k^{-\alpha}} = 1 \qquad (1.2)$$

Où ρ est la fonction d'autocorrélation et k le retard.

Par conséquent, les autocorrélations d'un processus à mémoire longue vérifient la relation asymptotique suivante :$\rho \sim c\, k^{-\alpha}$quand $c \to \infty$, où $c \in \mathbb{R}^+$ et $0 < \alpha < 1$.

1.2 Etude des processus ARIMA fractionnaire :

Dans le domaine des méthodes d'analyse des séries temporelles, il arrive souvent qu'on modélise des séries à mémoire longue au moyen de processus à mémoire courte tels que les modèles ARMA. Ceci revient alors à approximer la fonction d'autocorrélation qui décroît hyperboliquement au moyen d'une somme d'exponentielles. Même si une telle procédure est toujours possible, elle ne conduit pas à un modèle parcimonieux puisqu'il est nécessaire de considérer des retards très élevés dans la modélisation ARMA. Cette difficulté peut être résolue grâce à l'introduction des processus ARFIMA (AutoregressiveFractionnallyIntegratedMovingAverage) dont la caractéristique essentielle est la présence d'un paramètre d'intégration fractionnaire prenant explicitement en compte le comportement de long terme de la série.

Définition :

Un processus $\{X_t\}_{t\in Z}$ suit un processus ARFIMA (p,d,q) s'il satisfait l'équation suivante:

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t \qquad (1.3)$$

Où ε_t est un bruit blanc de variance σ^2, $\Phi(B)$ et $\Theta(B)$ sont deux polynômes caractéristiques d'ordre p et q respectivementdont les racines sont à l'extérieur du disque unité, et $d \in \mathbb{R}$.

$(1-B)^d$est appelé opérateur (Filtre) de différence fractionnaire.

$(1-B)^d$est le développement binomiale défini par:

$$(1-B)^d = \Delta^d = 1 - dB - \frac{d(1-d)}{2!}B^2 - \frac{d(1-d)(2-d)}{3!}B^3 - \cdots$$
$$= \sum_{j=0}^{+\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(1+j)}B^j$$

Où $\Gamma(.)$: correspond à la fonction Gamma.

On remarquera que les processus ARMA et ARIMA sont des cas particuliers des processus ARFIMA dans lesquels, respectivement $d = 0$ et dest un entier.

Propriétés :

Le processus ARFIMA le plus simple est le bruit fractionnaire, ou ARFIMA(0,d,0):

$$(1-B)^d X_t = \varepsilon_t \qquad (1.4)$$

Les principales propriétés de ce processus sont données par:

a) Lorsqued $< \frac{1}{2}$, $\{X_t\}_{t\in Z}$est un processus stationnaire et possède une représentation moyenne mobile infinie.
b) Lorsqued $> \frac{-1}{2}$, $\{X_t\}_{t\in Z}$est un processus inversible et a une représentation auto- régressive infinie.
c) Lorsqued $\geq \frac{1}{2}$, $\{X_t\}_{t\in Z}$est un processus non stationnaire.

Lorsque$-\frac{1}{2} < d < \frac{1}{2}$, le processus est stationnaire et inversible. La décroissance hyperbolique de la fonction d'auto corrélation indique

que les processus ARFIMA sont des processus à mémoire longue lorsque d est différent de zéro. En outre, le comportement de la densité spectrale aux basses fréquences montre que, lorsque d est positif, $\{X_t\}_{t\in Z}$est un processus persistant.

On peut alors établir une classification des séries temporelles d'après les résultats du théorème précédent en fonction des valeurs du paramètre d'intégration fractionnaire d:

- si $d = 0$, le processus ARFIMA (p,0,q) se réduit au processus ARMA standard et exhibe uniquement une mémoire de court terme (ne présente aucune structure de dépendance à long terme.

- Si $0 < d < \frac{1}{2}$le processus ARFIMA est un processus asymptotiquement stationnaire à mémoire longue. Les autocorrélations sont positives et diminuent hyperboliquement vers zéro lorsque le retard augmente. La densité spectrale est concentrée autour des faible fréquences (cycles lents), elle tend vers l'infini lorsque la fréquence tend vers zéro. On est face à un processus persistant.

- Si $-\frac{1}{2} < d < 0$ le processus est anti-persistant, les autocorrélations alternent de signe et la densité spectrale est dominée par des composantes de haute fréquence (la densité spectrale tend vers zéro lorsque la fréquence tend vers zéro).

2. Estimation par la méthode de Geweke Porter-Hudak (1983)

Cette méthode est l'une des méthodes d'estimation semi-paramétrique du paramètre d'intégration fractionnaire dd'un processus ARFIMA(p,d,q). Dans la suite, nous présentons cette méthode dite aussi log-périodogramme proposée par Geweke et Porter-Hudak (1983).

Principe de la méthode :

D'une manière générale, on considère un processus scalaire, $(X_t)_{t\in Z}$, faiblement

stationnaire, dont la densité spectrale est de la forme suivante sur l'intervalle [0,2π[:

$$f(\lambda) = \left|2\,sin\left(\frac{\lambda}{2}\right)\right|^{-2d} f^*(\lambda) \qquad (2.1)$$

Où

$$f^*(\lambda) = \frac{\sigma^2}{2\pi}\frac{\left|\theta(e^{-i\lambda})\right|^2}{|\phi(e^{-i\lambda})|^2} \qquad (2.2)$$

où d est le paramètre de mémoire compris dans l'intervalle $(-\frac{1}{2},\frac{1}{2})$ et f^* est une fonction continue bornée sur tout l'intervalle [0.2π[. le paramètre d contrôle le comportement de la densité spectrale dans un voisinage de zéro alors que f^* contrôle le comportement de courte mémoire.

En calculant l'équation (2.1) aux fréquences de Fourier : $\lambda_j = 2\pi_j/N$,

pour $j = 0, \dots, N-1$, où N est la taille de l'échantillon, et par passage aux logarithmes, on obtient alors:

$$log\, I_N(\lambda_j) = -2d\, log\left|2\,sin\left(\frac{\lambda_j}{2}\right)\right| + log\, f^*(0) + log\frac{f^*(\lambda_j)}{f^*(0)} + log\frac{I_N(\lambda_j)}{f(\lambda_j)} \qquad (2.3)$$

Où $I_N(\lambda_j)$ est le périodogramme (l'estimateur asymptotiquement sans biais de la densité spectrale (λ)) calculé à la fréquence λ_j, défini par l'expression:

$$I_N(\lambda) = \frac{1}{2\pi N}\left|\sum e^{i\lambda t}(X_t - \bar{X}_N)\right|^2 \qquad (2.4)$$

On considère alors l'équation (2.3) lorsque T tend vers l'infini, j étant fixé.

L'estimateur GPH nécessite à ce niveau deux hypothèses cruciales, relatives au comportement asymptotique des éléments de l'équation (2.3).

(H1) : pour des fréquences suffisamment basses ($\lambda \to 0$) le terme $log\frac{f^*(\lambda_j)}{f^*(0)}$ est négligeable.

(H2) : la suite des termes $log\frac{I_N(\lambda_j)}{f(\lambda_j)}$ pour $j = 1, \dots, m$ est asymptotiquement indépendante et identiquement distribuée (i.i.d) ; le nombre m de fréquences considérées est alors appelé la largeur de bande.

Sous les hypothèses (H1) et (H2), l'estimateur GPH du paramètre de mémoire d, noté $\hat{d}_{\text{GPH}}$, s'obtient alors en considérant la régression linéaire simple, pour $j = 1, \dots, m$:

$$log\, I_N(\lambda_j) = a + \beta Z_j + \varepsilon_j \qquad (2.5)$$

Où

- $a = log\, f^*(0) - \gamma$

- $Z_j = -2d\ log\left|2\ sin\left(\frac{\lambda_j}{2}\right)\right|$
- $\varepsilon_j = log\frac{I_N(\lambda_j)}{f(\lambda_j)}$

L'estimateur GPH est alors explicitement défini par l'égalité suivante :

$$\hat{d}_{GPH} = \frac{\sum_{j=1}^{m}(Z_j - \bar{Z})\ log\ I_N(\lambda_j)}{\sum_{j=1}^{m}(Z_j - \bar{Z})^2} \quad (2.6)$$

Geweke et Porter-Hudak(1983) montrent que, quand $-1/2 < d < 1/2$, la loi de l'estimateur, $\hat{d}_{GPH}$, de d tend vers une loi normale lorsque $N \to \infty$:

$$\hat{d}_{GPH} \sim N\left(d, \pi^2\left[6\sum_{j=1}^{m}(Z_j - \bar{Z})^2\right]^{-1}\right)$$

En suivant les suggestions de GPH (1983), le nombre de fréquence m est choisie de telle manière que m= N^{μ} , avec $\mu = 0.5, 0.6,\ 0.7$.

3. Expériences de simulation :

Dans le présent paragraphe, on effectuera une étude pratique des méthodes d'estimation de la mémoire longue à partir des simulations de Monté Carlo. On étudiera la technique de Geweke et Porter-Hudak en procédant au calcul de l'estimateur $\hat{d}_{GPH}$ pour plusieurs valeurs du nombre de fréquences m (m=T $^{\mu}$μ=0.4, 0.5, 0.6, sachant que T est la taille de l'échantillon), effectué par des simulations des différents processus ARFIMA. Tout d'abord on s'intéressera au cas ARFIMA(0,d,0) avec absence de mémoire courte, on effectue des

simulations à ces processus pour plusieurs valeurs de d (0.45, 0.30, …). On estime ces modèles et on calculera le biais, puis on procédera au test du coefficient d'intégration. Cette étape sera reproduite sur des processus mixte (présence de mémoire courte), on s'intéresse aux simulations des processus autorégressifs fractionnaires ARFIMA(1,d,0). Tous les résultats de simulation sont présentés dans des tableaux. Et après chaque simulation, les résultats seront interprétés.

3.1. Les hypothèses de travail:

Les simulations des processus à mémoire longue ARFIMA(0,d,0) et ARFIMA(0,d,1) sont faites à l'aide du package fracdiff du logiciel statistique R 2.6.1.

Les simulations de ces processus reposent sur certaines conditions et hypothèses:

1. Pour chaque expérience, différentes tailles d'échantillons ont été retenues: Petite (n = 100), moyenne (n = 500) et large (n = 750, n = 1500)
2. la variance de $\hat{d}_{GPH}$, peut être écrite par la formule:

$$var\left(\hat{d}_{GPH}\right) = \frac{\pi^2}{24m} + o\left(\frac{1}{m}\right) \quad avec \quad m = N^{\mu}, \quad \mu = 0.4, 0.5, 0.6$$

Elle dépend de la taille de l'échantillon et du nombre de fréquences. Par exemple pour toutes les séries de taille 300, et pour μ=0.5 la variance $var\left(\hat{d}_{GPH}\right) \approx 0.0237$et pour celles dont la taille est n=1500 la $var\left(\hat{d}_{GPH}\right) \approx 0.0106$.

3. Pour chaque cas, les résultats sont résumés dans des tableaux, où on trouve les vraies valeurs de d,de ϕ, et de θ(les coefficients autorégressifs et moyennes mobiles respectivement), la valeur estimée $\hat{d}_{GPH}$, la différence (vrai valeur – la valeur estimée) qui représente le biais moyen.

3.2. Les résultats des simulations par l'approche GPH:

a. Les processus ARFIMA(0,d,0):

Dans ce paragraphe, nous allons présenter les simulations réalisées sur des modèles de type ARFIMA pur : $(1-B)^d Y_t = \varepsilon_t$, ces modèles diffèrent selon plusieurs critères, à savoir la taille de l'échantillon, la valeur du paramètre d'intégration fractionnaire et la puissance utilisée pour déterminer le nombre de fréquences qui sont utilisées dans la régression qui permet d'estimer $\hat{d}_{GPH}$.

Les résultats de simulations dans le cadre de détection d'une mémoire longue dans les processus fractionnaires purs sont regroupés dans le tableau suivant.

D'après letableau de simulation n° 01, on voit bien que les estimateurs sont très proches des vraies valeurs pour toutes les tailles choisies. Et l'écart entre la valeur estimée et la vraie valeur ne dépasse pas 0.047, malgré les changements faites sur la taille d'échantillon et sur le nombre de fréquences m.

	d =	0,45			0,3			-0,2			-0,45		
N	μ =	0,4	0,5	0,6	0,4	0,5	0,6	0,4	0,5	0,6	0,4	0,5	0,6
1500	$\hat{d}_{GPH}$	0,489	0,459	0,464	0,319	0,317	0,302	-0,176	-0,189	-0,198	-0,381	-0,414	-0,469
	biais	-0,039	-0,009	-0,014	-0,019	-0,017	-0,002	-0,024	-0,011	-0,002	-0,069	-0,036	0,019
750	$\hat{d}_{GPH}$	0,473	0,468	0,476	0,279	0,295	0,286	-0,199	-0,186	-0,198	-0,419	-0,435	-0,468
	biais	-0,023	-0,018	-0,026	0,021	0,005	0,014	-0,001	-0,014	-0,002	-0,031	-0,015	0,018
300	$\hat{d}_{GPH}$	0,384	0,464	0,470	0,250	0,301	0,295	-0,233	-0,169	-0,191	-0,375	-0,448	-0,467
	biais	0,066	-0,014	-0,020	0,050	-0,001	0,006	0,033	-0,031	-0,009	-0,075	-0,002	0,017
100	$\hat{d}_{GPH}$	0,442	0,464	0,462	0,261	0,040	0,261	-0,153	-0,172	-0,189	-0,449	-0,439	-0,473
	biais	0,008	-0,014	-0,012	0,039	0,260	0,039	-0,047	-0,028	-0,011	-0,001	-0,011	0,023

Tableau 01: simulation des processus ARFIMA(0,d,0) pour les différents tailles d'échantillon N=150, N=750, N=300 et N=100

On conclut que le biais de l'estimateur $\hat{d}_{GPH}$ dans le cas des processus ARFIMA(0,d,0) est toujours petit, et l'estimateur $\hat{d}_{GPH}$ dans ce cas ne dépend ni de la taille d'échantillon, ni du nombre de fréquences utilisées.

b. Présence d'une mémoire courte:

Dans le paragraphe précédent, dans le traitement des processus fractionnaires purs ARFIMA(0,d,0) on s'est intéressé à l'effet de la taille et à l'effet de nombre des fréquences sur la précision de l'estimateur $\hat{d}_{GPH}$, de ce fait tous les cas cités avant (plusieurs taille et plusieurs nombres de fréquence) ont été traités.

Dans ce paragraphe, on va essayer d'étudier la robustesse de la méthode d'estimation de Geweke et Porter-Hudak sur les processus ARIMA fractionnaires avec une présence de mémoire courte, on

s'intéresse aux processus Auto Régressif AR(1), Moyenne Mobile MA(1), et ARFIMA (1,d,0). Puisque le but est d'étudier la présence de mémoire courte, on se contentera de trois tailles seulement, une petite (n=300), une moyenne (n=750) et une large (n=1500).

- **Les processus MA(1):**

Les résultats de l'estimateur du degré d'intégration $\hat{d}_{GPH}$ dans les processus simulés de type MA(1) : $Y_t = (1 - \theta B)\varepsilon_t$, sont représentés dans le tableau n°02.

	μ =	0,4				0,5				0,6			
N	θ =	0,7	0,3	-0,3	-0,7	0,7	0,3	-0,3	-0,7	0,7	0,3	-0,3	-0,7
1500	$\hat{d}_{GPH}$	**0,221**	0,032	0,004	0,015	**-0,433**	-0,033	0,006	0,012	**-0,183**	0,000	0,000	0,021
750	$\hat{d}_{GPH}$	-0,045	0,013	-0,005	-0,007	**-0,308**	-0,038	-0,004	0,010	**-0,341**	-0,012	0,019	-0,006
300	$\hat{d}_{GPH}$	-0,069	0,029	-0,069	0,029	**-0,253**	0,007	0,009	0,008	**-0,231**	0,006	0,005	-0,012

d'échantillon N=1500.

Tableau 02: simulation des processus MA(1) pour différents

D'après ce tableau, on remarque que le biais est faible dans tous les cas sauf dans les cas où les valeurs du paramètre moyenne mobile sont élevées et positives. Effectivement, si on observe le tableau dont μ=0.5et μ=0.6, on remarque que lorsque le paramètre θest proche de 1, le biais est important, alors que, par contre, pour les valeurs négatives de ce paramètre même les plus extrêmes, proche de -1, l'estimation est très appréciée. Et malgré la diminution de la taille d'échantillon, l'estimateur GPH de paramètre d'intégration fractionnaire est apprécié pour toutes les valeurs du paramètre moyenne mobile lorsque μ=0.4 sauf pour les valeurs élevés. Dans le cas où μ=0.6 et μ=0.5, le biais est élevé par apport à l'autre cas (μ=0.4), et le biais est toujours importantpour les valeurs élevées du paramètre moyenne mobile.

La remarque importante qu'on peut extraire de ce tableau, est que l'estimateur du paramètre d est sensible aux choix du nombre de

fréquence m par rapport à la taille de l'échantillon. Tel que, quand la taille de l'échantillon est petite, le paramètre $\hat{d}_{GPH}$ est plus précis lorsque le nombre de fréquence est petit. Donc le choix du nombre de fréquence à retenir dans la méthode dépend de la taille d'échantillon.

- **Les processus AR(1)**

Les résultats d'estimation du degré d'intégration $\hat{d}_{GPH}$ dans les processus simulés de type AR(1) : $(1-\varphi B)Y_t = \varepsilon_t$ sont représentés dans le tableau n° 03.

μ =	0,4				0,5				0,6			
φ =	0,7	0,3	-0,3	-0,7	0,7	0,3	-0,3	-0,7	0,7	0,3	-0,3	-0,7
$\hat{d}_{GPH}$	-0,022	0,016	0,008	-0,022	**0,433**	0,056	-0,019	-0,022	**0,093**	0,014	-0,008	-0,004
$\hat{d}_{GPH}$	0,012	0,030	-0,018	0,016	**0,433**	0,056	-0,008	-0,023	**0,211**	-0,047	-0,006	0,003
$\hat{d}_{GPH}$	0,005	-0,002	0,022	0,034	**0,273**	0,007	-0,008	-0,011	**0,221**	0,052	-0,005	-0,002

Tableau 03: simulation des processus AR(1) pour différents tailles d'échantillon N=1500, N=750 et N=300.

On remarque que les suggestions faites dans le paragraphe précédent (cas MA(1)) restent valables dans le cas des processus AR(1).

L'étude de l'estimateur $\hat{d}_{GPH}$ dans le cas des processus fractionnaires purs FI(d) a permis de constater que l'estimateur du degré d'intégration fractionnaire est robuste, et il ne dépend ni de la taille d'échantillon et ni du nombre de fréquence retenu dans l'estimation. La lecture du tableau 2 (respectivement 3) permet de tirer la même remarque dans le cas où les paramètres des processusMoyennes Mobiles (respectivement Autorégressifs) prennent des valeurs négatives. Pour les valeurs positives importantes du paramètre θ =0.7 (respφ =0.7) cette méthode (GPH) ne donne pas des meilleurs estimateurs dans tous les cas, tel que par exemple lorsque μ=0.6, le biais dépasse la valeur 0.34 lorsque N=750, et dépasse la valeur 0.18 lorsque n=1500. Donc on peut conclure que pour obtenir des meilleurs estimateurs par la

méthode GPH lorsque la taille de l'échantillon est petite et le paramètre de polynôme Autorégressive prend des valeurs positives, il faut choisir un nombre de fréquences m inférieur à $N^{0.5}$ oùN est la taille de l'échantillon.

c. Les processus mixtes

Les tableaux 04, 05 et 06 montrent les résultats de simulations effectuées pour l'estimation du degré d'intégration fractionnaire des processus ARFIMA (1,d,0): $(1-\varphi B)(1-B)^d Y_t = \varepsilon_t$.

d =	0,45				0,3			
φ =	0,7	0,3	-0,3	-0,7	0,7	0,3	-0,3	-0,7
	μ = 0,4				***μ = 0,4***			
$\hat{d}_{GPH}$	0,460	0,459	0,490	0,481	0,287	0,321	0,275	0,281
biais	-0,010	-0,009	-0,040	-0,031	0,013	-0,021	0,025	0,019
	μ = 0,5				***μ = 0,5***			
$\hat{d}_{GPH}$	0,491	0,458	0,443	0,464	0,346	0,317	0,319	0,284
biais	-0,041	-0,008	0,007	-0,014	-0,046	-0,017	-0,019	0,016
	μ = 0,6				***μ = 0,6***			
$\hat{d}_{GPH}$	0,547	0,477	0,459	0,450	0,390	0,317	0,313	0,290
biais	-0,097	-0,027	-0,009	0,000	-0,090	-0,017	-0,013	0,010

Tableau 04: simulation des processus ARFIMA(1,d,0) pour la taille

d =	0,45				0,3			
φ =	0,7	0,3	-0,3	-0,7	0,7	0,3	-0,3	-0,7
	μ = 0,4				***μ = 0,4***			
$\hat{d}_{GPH}$	0,499	0,479	0,470	0,429	0,318	0,286	0,278	0,312
biais	-0,049	-0,029	-0,020	0,021	-0,018	0,014	0,022	-0,012
	μ = 0,5				***μ = 0,5***			
$\hat{d}_{GPH}$	0,493	0,466	0,464	0,428	0,370	0,312	0,315	0,301
biais	-0,043	-0,016	-0,014	0,022	-0,070	-0,012	-0,015	-0,001
	μ = 0,6				***μ = 0,6***			
$\hat{d}_{GPH}$	0,499	0,467	0,467	0,464	0,449	0,334	0,279	0,295
biais	-0,049	-0,017	-0,017	-0,014	-0,149	-0,034	0,021	0,005

Tableau 05: simulation des processus ARFIMA(1,d,0) pour la taille d'échantillon N=750.

d =	0,45				0,3			
φ =	0,7	0,3	-0,3	-0,7	0,7	0,3	-0,3	-0,7
	μ = 0,4				***μ = 0,4***			
$\hat{d}_{GPH}$	0,499	0,445	0,435	0,498	0,367	0,309	0,228	0,283
biais	-0,049	0,005	0,015	-0,048	-0,067	-0,009	0,072	0,017
	μ = 0,5				***μ = 0,5***			
$\hat{d}_{GPH}$	0,499	0,499	0,422	0,437	0,337	0,315	0,322	0,308
biais	-0,049	-0,049	0,028	0,013	-0,037	-0,015	-0,022	-0,008
	μ = 0,6				***μ = 0,6***			
$\hat{d}_{GPH}$	0,499	0,498	0,425	0,424	0,394	0,342	0,317	0,297
biais	-0,049	-0,048	0,025	0,026	-0,094	-0,042	-0,017	0,003

Tableau 06: simulation des processus ARFIMA(1,d,0) pour la taille d'échantillon N=300.

De ces tableaux, on déduit les mêmes résultats que ceux de l'étude des processus AR(1) et MA(1). On voit bien que le biais est assez élevé lorsque la valeur du paramètre autorégressif est proche de 1 (à la limite de non stationnarité). Alors il existe un effet néfaste de la présence de mémoire courte sur l'estimateur $\hat{d}_{GPH}$.

Application Empirique :

Etude du phénomène ‘’Mémoire longue‘’ dans les données de température de l’air de la vile d’Alger

Dans ce paragraphe, nous présentons une application des processus à mémoire longue sur des données réelles. Et suite à l'introduction des modèles ARFIMA, beaucoup d'applications des processus à mémoire longue ont été développées dans la littérature statistique. En particulier, les domaines de la finance et de l'environnement constituent les champs d'applications de prédilection des chercheurs.

Notre applicationestorientée, principalement, dans une optique prévisionnelle. Ainsi, nous utilisons des critères de prévision comme éléments de comparaison entre les différents modèles estimés. Par ces applications nous mettons en évidence la capacité prédictive des processus à mémoire longue, en particulier sur des horizons de prévision de moyen et long terme. Cette application est réalisée sur des données climatiques relatives à la température de l'air dans la ville Alger.

Présentation des données:

Dans ce paragraphe, on s'intéresse à la série chronologique de type climatique: la série journalière de la température de l'air relevée à Alger, du premier janvier 2012 au 31 décembre 2013. Les températures sont en degrés Fahrenheit et la valeur de chaque jour correspond à la moyenne journalière des températures prise toute les 3 heures.

On note pour cette série par $(tempalg_t)_t$ dont une représentation graphique est donnée par la figure 01.

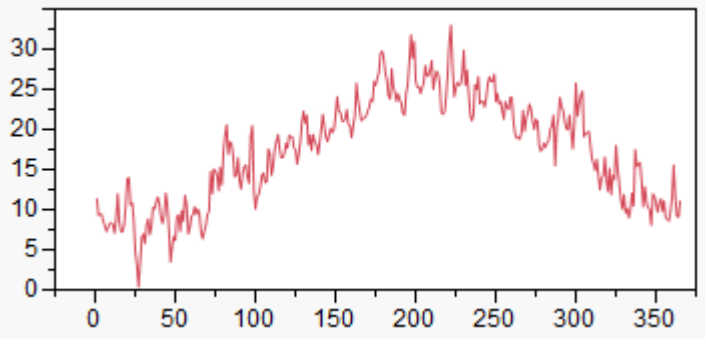

Figure 01: Evolution de la série journalière de température de l'air à Alger

Nous allons modéliser cette série en utilisant de manière compétitive deux approches différentes. La première approche est basée sur une modélisation linéaire à mémoire courte de type ARIMA, et la seconde approche est basée sur une modélisation à mémoire longue de type ARFIMA.De plus, pour chacune de ces deux approches, nous étudierons l'impact en prévision des différents modèles estimés.

Pour chaque modèle, on compare les capacités prédictives à l'aide du critère de la racine carré de l'erreur quadratique moyenne de prévision, noté RMSE ("RootMeanSquaredError"). Ce critère est défini comme suit:

$$\mathrm{RMSE} = \sqrt{\frac{1}{h}\sum_{i=1}^{h}(X_{t+i} - \widehat{X}_t(i))^2}$$

Où h est l'horizon de prévision et $\widehat{X}_t(i)$ est la valeur prédite de X_{t+i} Pour chaque approche, les prévisions sont calculées pour h=1, 3 et 6.

Première approche:

Dans cette première approche, nous utilisons la méthode classique de modélisation et de prévision de Box et Jenkins (1970). Et pour étudier la stationnarité de cette série on effectue le test de Dickey Fuller augmenté (ADF). Les résultats sont présentés dans le tableau suivant:

Modèles	**d**	coefficients	valeur	Stat observée	Probabilité
Modèle avec tendance et constante	0	ADF	-	-2.91951	0.1575
		Trend	0.00054	0.50906	0.6110
		C	0.76594	2.47546	0.0138
		Tempalg(-1)	-0.04960	-2.91951	0.0037
Modèle sans tendance avec constante	0	ADF	-	-0.93263	0.4261
		C	0.00131	0.74091	0.1064
		Tempalg(-1)	-0.04653	-0.93263	0.0036
Modèle sans tendance sans constante	0	ADF	-	-1.0882	0.5472
		Tempalg(-1)	-1.10347	-1.0821	0.0048
Modèle sans tendance sans constante	1	ADF	-	-21.1114	0.0000
		D(Tempalg(-1))	-1.10348	-21.1114	0.0000

Tableau 07: test ADF sur la série des températures de l'air de la ville d'Alger

D'après ce tableau on remarque que le coefficient de la tendance (modèle 01) est non significatif au seuil statistique 5%, et la constante n'est pas significative au même seuil (modèle 02), et la série n'est pas stationnaire dans le modèle 03où il n'y a ni tendance ni constante. Après la différentiation première de la série, on procède au test de racine unitaire. La valeur estimée de la statistique ADF (qui correspond à la t-statistique du coefficient de tempalg(-1)) est égale à -21.1114, cette valeur est inférieure à la valeur critique -1.95 au seuil statistique de 5% (modèle 04). En rejette en conséquence l'hypothèse nulle de présence d'une racine unitaire. Donc la série D(tempan(-1)) est stationnaire (intégrée d'ordre 1).

Après le passage par les quatre étapes de la méthodologie de Box et Jenkins, à savoir l'identification, l'estimation, la validation et la prévision, on retient le modèle ARIMA(1,1,1) pour la série $\Delta(\text{tempalg}_t)$ qui s'écrit comme suit :

$$(1 - 0.655\text{B})(1 - \text{B})(\text{tempalg}_t) = (1 - 0.885\text{B})\varepsilon_t$$

Les résultats relatifs à la qualité des prévisions obtenues à partir de ces deux modèles sont contenus dans le tableau 08.

Deuxième approche:

On s'intéresse maintenant à une approche longue mémoire de cette série journalière des températures de l'air, à l'aide d'un processus ARFIMA.

Dans un premier temps, on suppose que les ordres p et q des polynômes autorégressif et moyenne mobile sont nuls. On estime alors le paramètre de mémoire longue par la méthode GPH présentée précédemment, et nous prenons en considération la valeur de μ,(ici comme la taille de l'échantillon est petite n=365, nous prenons $\mu = 0.4$). On obtient le processus ARFIMA(0,d,0) suivant:

$$(1 - \mathrm{B})^{0.6167}\,\mathrm{tempalg}_t = \varepsilon_t$$

Dans une seconde étape, on cherche à spécifier correctement les ordres des parties, autorégressive et moyenne mobile de ces deux processus. Pour cela, on effectue une recherche de p et q (étape identification de la méthode Box et Jenkins), puis on passe à l'étape d'estimation des paramètres. On retient un processus de type ARMA(1,0) de la série qui s'écrit de la forme suivante:

$$(1 - 0.122\mathrm{B})(1 - \mathrm{B})^{0.6167}\,\mathrm{tempalg}_t = \varepsilon_t$$

On remarque d'après l'estimation que le paramètre de mémoire longue $\hat{d}$ estimé est superieur à 0.5, or l'intervalle du paramètre d'intégration fractionnaire d'un processus ARFIMA à mémoire longue est [-1/2,1/2[.

Dans ce cas-là, le processus considéré est alors non stationnaire. Donc on peut différencier la série, de manière à ce que le paramètre de mémoire longue soit dans $[-0.5,0.5[$. cette méthode considère uniquement le problème de l'estimation des paramètres, et elles ne se place pas d'un point de vus prévisionnel. Laurent FERRARA dans sa thèse de doctorat, a montré de manière empirique que le modèle qui correspond au processus ARFIMA non stationnaire, donne des prévisions qui convergent lentement vers la moyenne non conditionnelle du processus. Par contre les prévisions issues des

modèles ARFIMA des séries différencier ne converge pas. Donc, garder les données brutes, même en cas de non stationnarité, constitue une meilleure approche d'un point de vue de prévisionnelle.

On continue, alors, la procédure de prévision du modèle ARFIMA dans ce cas et voir sa qualité prédictive.

Les résultats relatifs à la qualité des prévisions sont dans le tableau suivant :

Modèle	Critère	Horizon		
		h = 1	h = 3	h = 6
ARIMA(1,1,1)	RMSE	1.3724	1.6077	1.7143
ARFIMA(1, 0.616, 0)		1.4117	1.4170	1.4170

Tableau 08: résultats relatifs à la qualité des prévisions de la série de température de l'air de la ville d'Alger à partir des processus ARIMA et ARFIMA.

Dans un but de comparaison des résultats de prévision des deux modèles choisies, on observe d'après le tableau 08, que sur un horizon de prévision de court terme (h = 1), le modèle à mémoire courte de la série $(tempalgt)_t$ est plus précis que le modèle ARFIMA. Par contre, lorsque l'horizon de prévision augmente (h = 3, h = 6), le modèle ARFIMA améliore leur performance, et il est plus précis que le modèle à mémoire courte.

Conclusion :

Dans ce papier nous avons concentré notre travail sur une méthode d'estimation du paramètre d'intégration fractionnaire,à savoir, la méthode de Geweke et Porter-Hudak.Ce travail a été fait essentiellement par une étude théorique basée sur des simulations des différents processus de type ARFIMA(p,d,q).

L'idée est d'étudier la convergence et la divergence de ces méthodes et leurs pouvoirs d'estimation. Pour cela, on a simulé plusieurs processus ARFIMA avec plusieurs valeurs de d, puis on est passé à l'estimation

par la méthode GPH en premier lieu pour plusieurs valeurs du nombre de fréquences m ($m=T^{\mu}$; μ=0.4, 0.5, 0.6). On a déduit,pour les résultats d'estimation des processus purs (ARFIMA(0,d,0)), que le biais de l'estimateur $\hat{d}_{GPH}$ est toujours petit, et l'estimateur $\hat{d}_{GPH}$ dans ce cas ne dépend ni de la taille d'échantillon, ni du nombre de fréquences utilisées. Et pour les processus mixtes (présence de mémoire courte) avec différentes valeurs du paramètre autorégressif (respectivement moyenne mobile), la première remarque qu'on peut souligner est le fort biais dans l'estimation du paramètre de longue mémoire des processus ARFIMA(1,d,0) et ARFIMA (0,d,1), en particulier lorsque la valeur du coefficient AR ou MA est proche de 1 ; ceci même sur des échantillons de grande taille. La deuxième remarque tirée de ces estimations est que l'estimateur du paramètre d est sensible aux choix du nombre de fréquence m par rapport à la taille de l'échantillon.

Ensuite nous avons présenté une application des processus à mémoire longue ARFIMA sur des données réelles de type climatique.L'étude de ces séries était orientée essentiellement dans une optique prévisionnelle, où on a comparé les capacités prédictives des processus à mémoire longue contre les processus à mémoire courte sur des horizons de prévision court, moyen et long terme.

Les résultats présentés sur les deux précédentes applications soulignent l'intérêt des processus ARFIMA, lorsqu'on désire effectuer des prévisions sur une série chronologique. En comparant avec les résultats obtenus en prévision par les processus à mémoire courte, les processus à mémoire longue sont performants sur un horizon moyen et long terme. Cependant, si on désire obtenir des prévisions à court terme (h=1 par example), il semble que les processus à mémoire courte soit plus efficace.

Références

Benyammi. Y : "Etude de l'estimateur de l'ordre d'integrationfractionnaire de geweke porter-hudak d'un processus ARFIMA et applications empiriques" thèse de magister, ENSSEA, Alger 2009.

Beran.J, Feng.Y, Ghosh.S et Kulik.R : "Long-Memory Processes : ProbabilisticProperties and StatisticalMethods", Springer-Verlag Berlin, 2013.

Beran.J, (1994) :"Statistics for long-memory processes. Monographs on statistics and appliedprobability" (Vol. 61). New York: Chapman and Hall/CRC.

Dickey, D.A. (1976) :"Estimation and hypothesistesting in nonstationary time series". PHD dissertation, Iowa State University.

Dickey, D.A. and W.A. Fuller (1981) :"Likelihood Ratio Statistics for Autoregressive Time Serieswith a Unit Root". Econometrica.

Ferrara.L: "Processus longue mémoire généralisés: Estimation, prévision et application". Thèse de doctorat, Paris 13, 01 Décembre 2000.

Geweke.J and Porter-Hudak.S: "The estimation and application of long memory time seriesmodels" Journal of time seriesanalysis Vol. 4. 1983.

Granger, C.W.J. and R. Joyeux (1980) :"An introduction to long memory time seriesmodels and fractionaldifferencing". J. Time SeriesAnalysis.

Hosking.J.R.M: "FractionalDifferencing" BIOMETRIKA , Vol. 68, No. 1. April 1981.

Hurvich, C.M., R.S. Deo and J. Brodsky (1998) :" The meansquarederror of Geweke and Porter-Hudakestimator of the memory paramater of a long memory time series". J. Time SeriesAnalysis.

Lardic.S et Mignon.V: "Prévision ARFIMA des taux de change: les modélisateurs doivent-ils encore exhorter à la naïveté des prévision?" Annales d'économie et de statistique, No 54, 1999.

Palma.W : "long-memory time series :Theory and methodes" , Wiley, canada 2007.