Les modèles de classe ARC

Hervé Alexandre, Marie-Claude Pichery

IME

INSTITUT DE MATHEMATIQUES ECONOMIQUES

LATEC C.N.R.S. URA 342

DOCUMENT de TRAVAIL

UNIVERSITE DE BOURGOGNE

FACULTE DE SCIENCE ECONOMIQUE ET DE GESTION

4, boulevard Gabriel - 21000 DIJON - Tél. 80 39 54 30 - Fax 80 39 56 48

ISSN : 0292-2002

n° 9301

LES MODELES DE CLASSE ARC

Hervé ALEXANDRE* - Marie-Claude PICHERY**

** Moniteur, LATEC, Faculté de Science Economique et de Gestion*
*** Professeur, LATEC, Faculté de Science Economique et de Gestion*

LES MODELES DE CLASSE ARCH

Résumé

La modélisation du processus ARCH a ouvert une nouvelle voie particulièrement riche pour la spécification des comportements sur les marchés financiers. A partir des travaux de ENGLE (1982), toute une série d'aménagements ont été proposés afin d'adapter à l'étude de cas particuliers les principes de cette modélisation. Ces divers modèles sont présentés dans ce document accompagnés des méthodes d'estimation et des formulations de tests appropriés.

Mots-clés

Processus ARCH, GARCH - méthode d'estimation - procédure de test.

Abstract

The aim of this document is a presentation of the Arch process which allows a new and powerful technic for modelling behaviour on financial markets. From the seminal paper of ENGLE (1982), numerous extensions were proposed to adapt this specification of the Arch process to particular situations. These models are presented here, with adapted estimation methods and some appropriate tests.

Keywords

Arch, Garch process, estimation method, tests.

L'étude des séries temporelles, particulièrement en finance, s'est trouvée confrontée à de nombreux problèmes, notamment ceux qui concernent la non-stationnarité et le caractère leptokurtique de la distribution des données observées. Plusieurs modèles ont été élaborés pour tenter de répondre à l'un ou l'autre de ces problèmes. On peut mentionner par exemple les travaux de MANDELBROT (1963) et FAMA (1965) sur l'utilisation des lois stables en finance. C'est cependant la modélisation du processus ARCH (**A**uto**R**egressive **C**onditionnal **H**eteroskedasticity), proposé par ENGLE en 1982 qui a ouvert une nouvelle voie particulièrement riche et fructueuse. Outre l'intérêt de l'expression de la kurtosis qui reflète plus fidèlement que les modèles classiques celle qui est constatée sur presque toutes les séries financières, un tel modèle intègre le caractère hétéroscédastique des processus à travers la variance conditionnelle. Une telle spécification permet de prendre en compte d'une façon élégante le fait que pour la plupart des séries financières la variabilité présente dépend du passé de façon plus ou moins importante. Il est ainsi possible de faire une description puis une prévision dynamique de la moyenne et de la volatilité d'une série.

Cette idée a été reprise et étendue à des formulations de plus en plus raffinées et nous nous proposons de présenter dans ce document les différents modèles élaborés et fondés sur l'hypothèse ARCH. La première section est consacrée à la présentation générale du modèle de base; elle est suivie de l'étude du modèle de régression ARCH (Section 2) et de l'exposé des procédures d'estimation et de prévision (Section 3). La Section 4 envisage une généralisation du processus à travers la modélisation GARCH, son estimation et le test de l'hypothèse. Enfin, la Section 5 fournit un ensemble d'extensions du modèle adaptées à différents problèmes économiques et financiers.

Section 1. Généralités.

Considérons le modèle suivant pour la période t :

$$y_t = x_t'\beta + \varepsilon_t$$

avec y_t : variable expliquée (scalaire)

x_t : vecteur des k variables explicatives

β : vecteur des k coefficients inconnus

ε_t : perturbation (terme aléatoire)

Il en résulte $\varepsilon = Y - X\beta$, pour l'ensemble de l'échantillon. Le processus étudié ici est celui de $\{\varepsilon_t\}$. Une hypothèse fréquente est que ε_t est un bruit blanc aux propriétés conditionnelles et non conditionnelles connues. Pour certaines séries économiques et en particulier financières, l'hypothèse de bruit blanc étant bien trop restrictive, de nouvelles hypothèses peuvent être envisagées telles que l'autorégressivité ou l'hétéroscédasticité.

§ 1. Autorégressivité.

Supposons que le processus $\{\varepsilon_t\}$ soit de forme autorégressive d'ordre 1 de telle sorte que $\varepsilon_t = \rho\ \varepsilon_{t-1} + u_t$ où u_t est un bruit blanc $(0, \sigma^2)$, indépendant de ε_{t-1} et $\rho < 1$. Les moments non conditionnels sont :

$$E(\varepsilon_t) = 0$$
$$V(\varepsilon_t) = \rho^2 V(\varepsilon_t) + \sigma^2 \qquad \text{d'où } V(\varepsilon_t) = \sigma^2/(1-\rho^2)$$

En termes conditionnels, on a :

$$E(\varepsilon_t/\varepsilon_{t-1}) = \rho\ \varepsilon_{t-1}$$
$$V(\varepsilon_t/\varepsilon_{t-1}) = \sigma^2$$

§2 Hétéroscédasticité.

A. Modèle classique.

Une des approches standard de l'hétéroscédasticité consiste à introduire une variable exogène z_t qui détermine la variance. Le modèle s'écrit alors $\varepsilon_t = u_t\, z_{t-1}$ où u_t est un bruit blanc $(0, \sigma^2)$ et z_{t-1} une variable non stochastique indépendante de u_t. Les propriétés non conditionnelles de ε_t sont alors :

$$E(\varepsilon_t) = 0$$
$$V(\varepsilon_t) = \sigma^2\, z_{t-1}^2$$

En termes conditionnels

$$E(\varepsilon_t/\varepsilon_{t-1}) = 0$$
$$V(\varepsilon_t / \varepsilon_{t-1}) = \sigma^2\, z_{t-1}^2$$

On constate que les propriétés des deux premiers moments, qu'ils soient conditionnels ou non, sont les mêmes ce qui limite l'intérêt de la formulation lorsqu'on cherche à prévoir des données futures, notamment dans le cas des cours boursiers.

B. Modèle bilinéaire (GRANGER et ANDERSEN, 1978).

GRANGER et ANDERSEN proposent une forme particulière d'hétéroscédasticité qui permet à la variance conditionnelle de ε_t de dépendre des réalisations passées de ε_t, réalisations qui vont constituer l'ensemble d'information I_t par rapport auquel la condition est définie. Le modèle se présente ainsi :

$\varepsilon_t = u_t \, \varepsilon_{t-1}$ où u_t est un bruit blanc $(0, \sigma^2)$ indépendant de ε_{t-1}

$E(\varepsilon_t) = 0$

$V(\varepsilon_t) = \sigma^2 V(\varepsilon_{t-1})$

En termes conditionnels on aura :

$E(\varepsilon_t / I_{t-1}) = 0$

$V(\varepsilon_t / I_{t-1}) = \sigma^2 \, \varepsilon_{t-1}^2$

On est ici en présence d'une formulation qui permet de distinguer entre la variance conditionnelle et non conditionnelle.

§3. Processus ARCH.

A. Définition du processus.

ENGLE (1982) introduit l'hétéroscédasticité à travers un schéma plus élaboré. Il pose $\varepsilon_t = u_t \, h_t$ où u_t est distribuée suivant une loi N(0,1) et supposée ici indépendante de h_t qui est une fonction de variables aléatoires donnée à travers son carré[1] :

$$h_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, \varepsilon_{t-i}^2$$

ε_t est appelé un processus ARCH d'ordre p dont les propriétés non conditionnelles sont les suivantes :

$E(\varepsilon_t) = E(u_t \, h_t) = E(u_t) \, E(h_t) = 0$

$V(\varepsilon_t) = \sigma^2$, quantité constante dont l'expression en terme des α_i sera donnée plus loin.

En termes conditionnels il vient :

$E(\varepsilon_t / I_{t-1}) = E(u_t / I_{t-1}) \, E(h_t / I_{t-1}) = 0$

$V(\varepsilon_t / I_{t-1}) = h_t^2$ quantité variable dans le temps.

[1]En utilisant la notation des polynômes scalaires de retard, nous adopterons l'écriture suivante :
$h_t^2 = \alpha_0 + \tilde{\alpha}(L) \, \varepsilon_t^2$ avec $\tilde{\alpha}(L) = \alpha_1 L + \ldots + \alpha_p L^p$

Une telle formulation permet d'introduire dans un modèle une variance dont les changements sont endogènes[2]. Cette notion a été pressentie auparavant et se révèle intéressante en finance. En effet MAC NEES, en 1979, constate, d'une part que *l'incertitude liée à la prévision varie selon les périodes et pas seulement avec l'horizon de prévision* et d'autre part que *les erreurs se regroupent souvent en erreurs élevées suivies (et précédées) d'erreurs plus faibles*. MANDELBROT (1963) avait relevé ce phénomène et prônait l'utilisation des lois de LEVY-PARETO stables qui peuvent posséder une variance théorique infinie mais qui sont difficilement utilisables en pratique. Comme par ailleurs l'inadaptation de la loi normale en finance a été mise en évidence par FAMA (1965 et 1970) et MUSSEN (1979), il était nécessaire de proposer une formulation apte à représenter de telles évolutions de la variance. C'est avec le processus ARCH qu'il est possible de prendre totalement en compte l'hétéroscédasticité et en particulier la variabilité de la variance.

A. Condition d'existence des moments d'ordre r.

Le processus est défini par $\varepsilon_t / I_{t-1} \sim N(0, h_t^2)$. ENGLE donne les conditions d'existence des moments non conditionnels pairs[3] par le théorème suivant (les moments impairs sont nuls) :

> Théorème 1.
>
> Pour un entier r, le 2r^{ème} moment d'un processus ARCH d'ordre 1 existe si :
>
> $$\alpha_1 \prod_{j=1}^{r} (2j-1) < 1 \qquad \text{avec } \alpha_0 > 0 \text{ et } \alpha_1 \geq 0$$

B. Stationnarité du processus.

Comme pour tout processus stochastique, il faut définir les conditions de stationnarité. Un processus ARCH d'ordre p avec α_0 positif et $\alpha_1,\ldots,\alpha_p$ positifs ou nuls sera stationnaire au deuxième ordre si et

---

[2] Si p = 0, on retrouve un bruit blanc gaussien.

[3] Si les conditions d'existence de la variance non conditionnelle d'un processus ARCH ne sont pas satisfaites, il faut alors envisager l'emploi des lois stables de LEVY.

seulement si l'équation caractéristique associée ($\tilde{\alpha}(L)=0$) a toutes ses racines hors du cercle unitaire. La variance non conditionnelle est constante et égale à :

$$E(\varepsilon_t^2) = \sigma^2 = \frac{\alpha_0}{1-\tilde{\alpha}(1)}$$

C. Kurtosis d'un processus ARCH.

La kurtosis est le rapport du moment centré d'ordre 4 au carré du moment centré d'ordre 2. Pour le processus ARCH(1) suivant :

$$\varepsilon_t / I_{t-1} \sim N\left(0\,,\,\alpha_0 + \alpha_1\,\varepsilon_{t-1}^2\right)$$

et sous la condition d'existence de ces moments, la kurtosis est égale à:

$$k = \frac{E\left(\varepsilon_t^4\right)}{E\left(\varepsilon_t^2\right)^2} = \frac{\dfrac{3\alpha_0^2}{\left(1-\alpha_1\right)^2}\cdot\dfrac{1-\alpha_1^2}{1-3\alpha_1^2}}{\left(\dfrac{\alpha_0}{1-\alpha_1}\right)^2} = \frac{3\left(1-\alpha_1^2\right)}{1-3\alpha_1^2}$$

Cette expression fait apparaître que k est toujours supérieur à 3 qui est la kurtosis de la loi normale. Un processus ARCH a donc toujours une distribution leptokurtique (la série possède des queues de distributions plus épaisses que celle d'une loi normale). L'utilisation de ce processus dans l'étude des séries financières va amener un plus grand réalisme en tenant mieux compte des événements rares.

Section 2. Modèle de régression ARCH.

Le modèle de régression ARCH est donné par la spécification suivante :

$$y_t = x_t'\beta + \varepsilon_t \qquad (1)$$

$$\varepsilon_t \sim \text{loi}\,(0,\sigma^2)$$

$$\varepsilon_t / I_{t-1} \sim N(0, h_t^2)$$

$$y_t/I_{t-1} \sim N(x_t'\beta, h_t^2)$$
$$h_t^2 = \alpha_0 + \tilde{\alpha}(L)\varepsilon_t^2 \qquad (2)$$

où les paramètres inconnus sont regroupés dans les vecteurs α (d'ordre p+1) et β (d'ordre k). L'expression (2) qui correspond à une variance, doit être strictement positive, ce qui est imposé par le théorème 1 (qui fournit des conditions suffisantes).

§1. Identification de l'ordre du processus.

En utilisant l'approximation $\varepsilon_t^2 = h_t^2 + v_t$ où v_t est un bruit blanc, la variance conditionnelle s'exprime comme un processus AR(p) sur $\{\varepsilon_t^2\}$ avec un terme constant (BOLLERSLEV,1986). Le choix des retards (p) dans l'expression de la variance conditionnelle n'est pas arbitraire et se déduit par le biais de l'analyse de BOX et JENKINS (1970). La valeur de p est obtenue à partir de la fonction d'autocorrélation partielle calculée sur la série des carrés des résidus de la régression de Y sur X.

La fonction d'autocovariance est:

$$\gamma_n = \gamma_{-n} = Cov(\varepsilon_t^2, \varepsilon_{t-n}^2)$$

La fonction d'autocorrélation en est déduite :

$$\rho_n = \frac{\gamma_n}{\gamma_0} = \frac{Cov(\varepsilon_t^2, \varepsilon_{t-n}^2)}{Cov(\varepsilon_t^2, \varepsilon_t^2)} = \frac{Cov(\varepsilon_t^2, \varepsilon_{t-n}^2)}{V(\varepsilon_t^2)}$$

Les autocorrélations partielles sont déduites pas à pas des équations de YULE-WALKER. En pratique r_n (autocorrélation partielle) et ρ_n sont inconnues et seront estimés par $\widehat{r_n}$ et $\widehat{\rho_r}$ calculées sur la base de l'échantillon. Ce sont des estimateurs convergents en probabilité des autocorrélations théoriques (GRANGER et NEWBOLD (1977), MONFORT et GOURIEROUX, (1990 a), p 374).

§2. Méthode d'estimation.

Les estimateurs de α et β obtenus par la méthode des Moindres Carrés Ordinaires ne sont pas BLUE. Il est donc nécessaire d'employer la méthode du Maximum de Vraisemblance ou la méthode des Moindres Carrés

Généralisés. La première va être développée ici; la seconde est donnée ultérieurement en raison de sa commodité d'emploi, mais elle ne fournit pas des estimateurs asymptotiquement efficients.

Un premier problème apparaît du fait que la distribution (non conditionnelle) du terme stochastique ε_t est inconnue et que par conséquent, la fonction de vraisemblance ne peut être construite. Lorsque T est grand, il est possible d'approximer la log-vraisemblance par la log-vraisemblance conditionnelle (MONFORT et GOURIEROUX, 1990 a, pp 383-384). En effet, la loi peut être explicitée par le biais des diverses lois conditionnelles successives. L'estimation des paramètres inconnus α et β va donc être estimée sous l'hypothèse de normalité de la distribution conditionnelle des erreurs. Les estimateurs obtenus seront ceux du pseudo maximum de vraisemblance.

Etant donné la fonction de densité conditionnelle pour la période t et son logarithme

$$f_{x,I}(\varepsilon_t/I_{t-1}) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{h_t^2}} \cdot \exp\left[- \frac{1}{2h_t^2}\left(y_t - x_t' b \right)^2 \right]$$

$$l_t = \ln\left(f(\varepsilon_t/I_{t-1})\right) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_t^2) - \frac{\varepsilon_t^2}{2h_t^2}$$

Sous l'hypothèse d'indépendance des erreurs, on définit la fonction de pseudo vraisemblance :

$$L = \sum_{t=1}^{T} l_t = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(h_t^2) - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2h_t^2}$$

Cette fonction L doit être maximisée par rapport aux coefficients des vecteurs α et β.

Les conditions de premier ordre sont :

$$\frac{\delta L}{\delta \alpha} = \sum_{t=1}^{T} \frac{1}{2h_t^2}\left[\frac{\varepsilon_t^2}{h_t^2} - 1\right]\frac{\delta h_t^2}{\delta \alpha} = 0 \qquad (1)$$

$$\frac{\delta L}{\delta \beta} = \sum_{t=1}^{T} \frac{\varepsilon_t x_t}{h_t^2} - \sum_{t=1}^{T}\frac{1}{h_t^2}\left[\frac{\varepsilon_t^2}{h_t^2} - 1\right]\sum_{i=1}^{p}\alpha_i \varepsilon_{t-i} x_i = 0 \qquad (2)$$

Sachant que $z_t = \frac{\delta h_t^2}{\delta\alpha} = \begin{pmatrix} 1 \\ \varepsilon_{t-1}^2 \\ \dots \\ \varepsilon_{t-p}^2 \end{pmatrix}$, la première série de conditions

permet d'établir que $\sum_{t=1}^{T} \frac{1}{2h_t^2}\left[\frac{\varepsilon_t^2}{h_t^2} - 1\right] = 0$. L'ensemble des relations constitue un système d'équations non linéaires par rapport aux paramètres; elles doivent être résolues par des méthodes numériques.

§3. Estimation par le pseudo maximum de vraisemblance.

L'estimation paraît assez ardue et semble devoir se faire simultanément pour α et β. ENGLE démontre que l'on peut procéder à une estimation séparée s'il y a indépendance entre chaque estimateur ce qui se traduit par la nullité des termes non diagonaux de la matrice d'information (matrice bloc). Nous allons donc déterminer cette matrice puis les conditions sous lesquelles les vecteurs α et β sont indépendants.

A. Matrice d'information.

La matrice d'information est la matrice symétrique d'ordre (p+1+k) construite à partir des dérivées partielles secondes de la log-vraisemblance L.

$$\begin{bmatrix} R_{\alpha\alpha} & R_{\alpha\beta} \\ R_{\beta\alpha} & R_{\beta\beta} \end{bmatrix}$$

Les dérivées secondes sont données par (ENGLE, 1982) :

$$\frac{\delta^2 L}{\delta\alpha\delta\alpha'} = -\sum_{t=1}^{T} \frac{1}{2h_t^4} \frac{\delta h_t^2}{\delta\alpha} \frac{\delta h_t^2}{\delta\alpha'} \frac{\varepsilon_t^2}{h_t^2} - \sum_{t=1}^{T} \left[\frac{\varepsilon_t^2}{h_t^2} - 1\right] \frac{1}{2h_t^4} \frac{\delta h_t^2}{\delta\alpha} \frac{\delta h_t^2}{\delta\alpha'}$$

$$\frac{\delta^2 L}{\delta\beta\delta\beta'} = \sum_{t=1}^{T} \frac{x_t x_t'}{h_t^2} - \sum_{t=1}^{T} \frac{1}{2h_t^4}\,\frac{\delta h_t^2}{\delta\beta}\,\frac{\delta h_t^2}{\delta\beta'}\,\frac{\varepsilon_t^2}{h_t^2}$$

$$- 2\sum_{t=1}^{T} \frac{\delta h_t^2}{\delta\beta}\,\frac{\varepsilon_t x_t'}{h_t^2} + \sum_{t=1}^{T}\left[\frac{\varepsilon_t^2}{h_t^2} - 1\right]\frac{\delta}{\delta\beta'}\left[\frac{1}{2\,h_t^2}\,\frac{\delta h_t^2}{\delta\beta}\right]$$

On déduit la matrice d'information de ce qui précède sachant que les espérances conditionnelles vérifient :

$$E\left[\frac{\varepsilon_t^2}{h_t^2} \;/\; I_{t-1}\right] = 1 \qquad\qquad E\left[\frac{\varepsilon_t^2}{h_t^2} - 1 \;/\; I_{t-1}\right] = 0$$

Pour le bloc $R_{\alpha\alpha}$ et son estimation $\hat{R}_{\alpha\alpha}$:

$$R_{\alpha\alpha} = \sum_{t=1}^{T} E\left[\frac{1}{2h_t^4}\,\frac{\delta h_t^2}{\delta\alpha}\,\frac{\delta h_t^2}{\delta\alpha'}\right] \qquad\qquad \hat{R}_{\alpha\alpha} = \sum_{t=1}^{T}\left[\frac{1}{2h_t^4}\,\frac{\delta h_t^2}{\delta\alpha}\,\frac{\delta h_t^2}{\delta\alpha'}\right]$$

Pour le bloc $R_{\beta\beta}$ et son estimation $\hat{R}_{\beta\beta}$:

$$R_{\beta\beta} = \sum_{t=1}^{T} E\left[\frac{x_t x_t'}{h_t^2} + \frac{1}{h_t^4}\,\frac{\delta h_t^2}{\delta\beta}\,\frac{\delta h_t^2}{\delta\beta'}\right] \qquad\qquad \hat{R}_{\beta\beta} = \sum_{t=1}^{T}\left[\frac{x_t x_t'}{h_t^2} + 2\sum_{j=1}^{p} \alpha_j^2\frac{\varepsilon_{t-j}^2}{h_t^4}\,x_{t-j}x_{t-j}'\right]$$

Pour le bloc $R_{\alpha\beta}$ et son estimation $\hat{R}_{\alpha\beta}$:

$$R_{\alpha\beta} = \sum_{t=1}^{T} E\left[\frac{1}{2h_t^4}\,\frac{\delta h_t^2}{\delta\alpha}\,\frac{\delta h_t^2}{\delta\beta'}\right] \qquad\qquad \hat{R}_{\alpha\beta} = \sum_{t=1}^{T}\left[\frac{1}{2h_t^4}\,\frac{\delta h_t^2}{\delta\alpha}\,\frac{\delta h_t^2}{\delta\beta'}\right]$$

Nous devons maintenant nous attacher à trouver les conditions de nullité de $R_{\alpha\beta}$ afin de pouvoir estimer séparément α et β sans perte d'efficience. ENGLE (1982, p 996) établit un théorème qui fournit seulement une condition suffisante qui repose sur les conditions de symétrie et de régularité d'un processus ARCH.

Théorème 2.

Si un processus ARCH est symétrique et régulier alors $R_{\alpha\beta} = 0$

B. Symétrie et régularité d'un processus ARCH.

1. Processus ARCH symétrique.

En utilisant la convention d'écriture ci-dessous[4], soit A_t un vecteur aléatoire de dimension p extrait de la population Θ qui a comme élément $A_t = (A_{t-1},\dots,A_{t-p})$. Pour chaque A_t construisons A^*_{tm} identique à A_t sauf le $m^{ième}$ élément qui est multiplié par -1, pour $m = 1,\dots,p$.

Définition 1.

Un processus ARCH d'ordre p est symétrique si :

(a) $h^2(A_t) = h^2(A^*_{tm}) \qquad \forall A_t \in \theta$

(b) $\delta h^2(A_t)/\delta\alpha_i = \delta h^2(A^*_{tm})/\delta\alpha_i \qquad \forall i,m = 1,\dots,p$

(c) $\delta h^2(A_t)/\delta A_{t-m} = \delta h^2(A^*_{tm})/\delta A^*_{t-m} \qquad \forall m = 1,\dots,p$

2. Processus ARCH régulier.

Définition 2.

Le processus ARCH d'ordre p est régulier si :

(a) $\min h^2_t(A_t) \le \zeta \qquad \forall \zeta > 0$ et $\forall A_t$

(b) $E\left|\delta h^2(A_t)/\delta\alpha_i\right|\left|\delta h^2(A_t)/\delta A_{t-m}\right|/I_{t-m-1}$ existe $\forall$ i, t, m

La première condition est celle de positivité de la variance conditionnelle. Nous utiliserons de préférence comme condition de régularité du processus, le théorème 3 donné par ENGLE (1982) :

Théorème 3.

Un processus ARCH d'ordre p satisfait à la condition de régularité si $\alpha_0 > 0$

et $\alpha_1,\dots,\alpha_p \ge 0$.

---

4 $h^2(\varepsilon_t) = \alpha_0 + \tilde{\alpha}(L)\,\varepsilon_t^2$, conformément à l'écriture donnée dans la note 1

Section 3. Procédure d'estimation et prévision.

Nous présentons ici l'algorithme d'estimation d'un modèle de régression ARCH ; mais auparavant il semble utile d'expliciter le test ARCH qui permet d'éviter de s'engager dans l'estimation en vain.

§1. Test ARCH.

La complexité de la méthode d'estimation incite à s'interroger sur l'existence de l'hypothèse ARCH. C'est pourquoi avant de commencer l'estimation, un test de cette hypothèse est réalisé. C'est le test du multiplicateur de Lagrange qui semble le plus approprié (cf BREUSCH et PAGAN 1978, GODFREY 1978 ou ENGLE 1979).

On teste l'hypothèse nulle :

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

contre l'hypothèse alternative :

$$H_a: \alpha_i \text{ non tous nuls } (0 < i \leq p)$$

Considérons le modèle ARCH avec

$$h_t^2 = h(z_t'\alpha) \quad \text{où} \quad z_t' = (1, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-p}^2)$$

h est une fonction quelconque et les ε_t sont les erreurs issues de la régression $Y = X\beta + \varepsilon$ par les MCO. Sous l'hypothèse nulle, h_t^2 est réduite à la constante $h° = \alpha_0$.

Posons

$$\frac{\delta h_t^2}{\delta \alpha} = h'(z_t' \alpha) = h'$$

Dans le cas général et sous l'hypothèse nulle, le score (vecteur d'ordre p+1) et la matrice d'information s'écrivent sous la forme générale :

$$\frac{dL^0}{d\alpha} = \frac{h'^0}{2h^0} \sum_{t-1}^{T} z_t' \left[\frac{\varepsilon_t^2}{h^0} - 1\right] = \frac{h'^0}{2h^0} Z' f^0$$

$$R_{\alpha\alpha}^0 = \frac{1}{2}\left[\frac{h'^0}{h^0}\right]^2 E\,(Z'Z)$$

où $Z' = (z_1 \dots z_T)$ est la matrice d'ordre $(p+1, T)$

h'° est la dérivée de h sous l'hypothèse nulle (ici il s'agit d'un scalaire)

f° est le vecteur colonne (T,1) de terme générique $\left[\frac{\varepsilon_t^2}{h^o} - 1\right]$

Dans le problème traité ici, la fonction $\mathbf{h}_t^2$ s'écrit $z_t'\alpha$ et sa dérivée n'est rien d'autre que le vecteur z_t d'ordre p+1. Le score et la matrice d'information s'écrivent de la manière suivante sous l'hypothèse nulle :

$$\frac{dL^0}{d\alpha} = \frac{1}{2\,\alpha_0} Z'\, f^0 \qquad R^o_{\alpha\alpha} = \frac{1}{2\alpha_0^2} E\,(Z'Z)$$

La statistique du multiplicateur de Lagrange prend la forme :

$$L^* = {}^1/_2\ f^{\circ\prime}Z(Z'Z)^{-1}Z'f^{\circ}$$

Sachant que sous l'hypothèse de normalité, plim $f^{\circ\prime}f^{\circ}/T = 2$, on arrive à un test asymptotique équivalent, construit à partir de la statistique :

$$L = {}^{(2T\ L^*)}/_{(f^{\circ\prime}f^{\circ})} = T\ R^2$$

où R^2 est le coefficient de corrélation multiple de la régression f° sur Z.

Le test d'un processus ARCH(p) s'effectue alors en quatre étapes :

1) régression de Y sur X par les moindres carrés ordinaires;
2) calcul des résidus ε_t, puis de leur carré;
3) régression de ces carrés sur une constante et p retards

$$\varepsilon_t^2 = \alpha_0 + \sum_{i-1}^{p} \alpha_i \varepsilon_{t-i}^2$$

4) test de la statistique L qui est distribuée selon un X^2 à p degrés de liberté.

Cette procédure fut employée par GRANGER et ANDERSON (1978) pour tester différentes hypothèses sur les moments des séries bilinéaires.

§2 Algorithme du score et initialisation.

Une fois l'hypothèse ARCH(p) retenue et sous l'hypothèse que la matrice d'information est bloc diagonale (lorsque la condition suffisante de symétrie et de régularité du processus est vérifiée), il est possible d'estimer α et β séparément sans perte d'efficience asymptotique. La variance σ^2 est estimée par $\alpha_0/(1-\tilde{\alpha}(1))$ et obtenue à partir des paramètres calculés à la dernière itération de l'algorithme présenté ci-dessous.

Soit Φ le vecteur des paramètres à estimer. La procédure du score est une procédure itérative pour laquelle on aura à l'étape i+1 :

$$\Phi^{i+1} = \Phi^i + \left[\hat{R}^i_{\Phi\Phi}\right]^{-1} \frac{1}{T} \sum_{t=1}^{T} \frac{\delta l_t^i}{\delta \Phi}$$

où $\hat{R}^i_{\Phi\Phi}$ et $\frac{\delta l_t^i}{\delta \Phi}$ sont respectivement la matrice d'information et le score estimés à l'étape i. Une valeur initiale doit être fixée pour lancer le processus; il s'agit du vecteur β estimé par les Moindres Carrés Ordinaires appliqués au modèle $Y = X\beta + \varepsilon$; Les résidus de cette régression sont ensuite utilisés pour estimer le paramètre α.

A. Estimation de α.

$$\alpha^{i+1} = \alpha^i + (w'w)^{-1} w' f^i$$

où $W' = (w'_1, \ldots, w'_T)$ avec $w_t = \left(1, \varepsilon^2_{t-1}, \ldots, \varepsilon^2_{t-p}\right) / h_t^{2i}$

$f^{i\prime} = (f_1^i, \ldots, f_T^i)$ avec $f_t^i = \left(1, \varepsilon_t^2 - h_t^{2i}\right) / h_t^{2i}$

ε_t : résidu de l'itération i

h_t^{2i} : variance conditionnelle à l'étape i

α^i : vecteur des α_j (j = 1,...,p) estimé à l'itération i

L'expression laisse apparaître la possibilité d'estimer le vecteur α à chaque étape à partir d'une régression de f^i sur W par les MCO.

B. Estimation de β.

Pour une estimation donnée de α, l'estimation de β est obtenue à partir d'une procédure itérative basée sur la relation :

$$\beta^{i+1} = \beta^i + \left(\overline{X}'\overline{X}\right)^{-1} \overline{X}'\overline{e}$$

dans laquelle

$$\bar{X}_t = X_t r_t \qquad \text{avec} \qquad r_t = \sqrt{h_t^{-2} + 2\varepsilon_t^2 \sum_{j-1}^{p} \alpha_j^2 h_{t+j}^{-4}}$$

$$\bar{e}_t = e_t\left(\frac{s_t}{r_t}\right) \qquad \text{avec} \qquad s_t = h_t^{-2} - \sum_{j-1}^{p} \alpha_j^2 h_{t+j}^{-4}\left(\varepsilon_{t+j}^2 - h_{t+j}^2\right)$$

Comme précédemment, à chaque étape on obtient β à partir du précédent en ajoutant le résultat de la régression de $\bar{e}$ (vecteur T,1) sur $\bar{X}$ (matrice T,k) par les MCO.

C. Propriétés des estimateurs.

Pour un vecteur Φ de paramètres à estimer, BELSLEY (1979) définit un critère de convergence construit à partir du scalaire suivant :

$$\theta = \left(\frac{\delta L}{\delta \Phi}\right)' \frac{\delta^2 L}{\delta \Phi \delta \Phi'} \frac{\delta L}{\delta \Phi}$$

Ce terme θ peut s'interpréter comme la partie résiduelle dans un développement de TAYLOR de l'expression à maximiser. La procédure algorithmique s'arrête lorsque θ est inférieur à une valeur fixée au début de la procédure.

Par le théorème de CROWDER (1976), on démontre que les distributions asymptotiques de α_{MV} et de β_{MV} sont données par :

$$\sqrt{T}(\hat{\alpha} - \alpha) \sim N(0, R_{\alpha\alpha}^{-1})$$

$$\sqrt{T}(\hat{\beta} - \beta) \sim N(0, R_{\beta\beta}^{-1})$$

Une estimation des matrices de variances-covariances asymptotiques des paramètres, calculée à la dernière itération, est donnée par

$$V(\hat{\alpha}) = \hat{\sigma}^2 (W'W)^{-1} \qquad V(\hat{\beta}) = \hat{\sigma}^2 (\bar{X}'\bar{X})^{-1} \qquad \text{avec} \qquad \hat{\sigma}^2 = \hat{\alpha}_0 \Big/ \left(1 - \tilde{\alpha}(1)\right)$$

§3. Une méthode alternative : les M.C.G. réalisables.

Le problème posé par les M.C.O. (estimateurs non efficients) pousse à rechercher d'autres méthodes d'estimation. Le maximum de vraisemblance semble la meilleure méthode mais les paramètres peuvent également être estimés par les Moindres Carrés Quasi Généralisés (ZAKOIAN, 1990) ou les Moindres Carrés Généralisés réalisables selon la terminologie de AITKEN.

Soit le modèle dont l'expression en t est

$$Y_t = x'_t\beta + \varepsilon_t \qquad (1)$$

A. Première étape.

1. Estimation de β.

En appliquant les M.C.O. sur l'équation (1) nous obtenons une estimation centrée de β. Nous pouvons alors calculer les résidus estimés.

2. Estimation de α.

Si ε_t suit un processus ARCH(p) alors ε_t^2 peut être décrit comme suivant un AR(p) avec constante, donc:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i\varepsilon_{t-i}^2 + u_t$$

Il suffit alors pour estimer α, de régresser ε_t^2 par les M.C.O. sur une constante et p retards.

B. Deuxième étape.

1. Estimation de β.

La variance (non conditionnelle) de ε_t peut être approchée par h_t^2 estimée précédemment par:

$$\hat{h}_t^2 = \hat{\alpha}_0 + \sum_{i-1}^{p} \hat{\alpha}_i\hat{\varepsilon}_{t-i}^2$$

Il est possible alors d'obtenir une nouvelle estimation de β en utilisant la méthode des M.C.Q.G. (Moindres Carrés Quasi Généralisés) et une estimation de la matrice de variances-covariances $\Omega = \mathrm{diag}(h_t^2)$. L'estimateur de β ainsi calculé admet la distribution asymptotique suivante :

$$\sqrt{T}\left(\hat{\beta}_{MCG} - \beta\right) \sim N\left(0\,,\,\sigma^2\left(X'\Omega^{-1}X\right)^{-1}\right)$$

Si on compare cette variance asymptotique à $R_{\beta\beta}$, on constate que β_{MCG} n'est pas asymptotiquement efficient.

2. Estimation de α.

En ce qui concerne le vecteur α,on a :

$$V\left(\varepsilon_t^2 / \varepsilon_{t-1}\right) = 2h_t^2 \text{ estimée par } 2\hat{h}_t^2 \qquad (2)$$

A la seconde étape, les estimateurs de α_i (i=0, ..., p) sont obtenus en régressant ε_t^2 sur une constante et $\varepsilon_{t-1}^2 \ldots \varepsilon_{t-p}^2$ par MCG réalisables et en utilisant la structure de variance-covariance définie par (2)

§4. La prévision.

Soit le modèle de régression dont les résidus sont engendrés par un processus ARCH(p); pour la période t, le modèle est le suivant :

$$Y_t = x'_t\beta + \varepsilon_t \qquad \varepsilon_t \sim NID(0,\sigma^2) \qquad \varepsilon_t / I_{t-1} \sim N(0,h_t^2)$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2$$

A la période T, une prévision de Y à l'horizon k est notée $Y_T^*(k)$; sous l'hypothèse de normalité des erreurs, la meilleure prévision (au sens de l'erreur quadratique moyenne minimale) est une fonction linéaire des erreurs présente et passées.

$$Y_T^*(k) = f(\varepsilon_{T-j};\ j = 0, \ldots, r) = \sum_{j=0}^{r} b_j\, \varepsilon_{T-j}$$

L'erreur de prévision, $e_T(k)$, différence entre la valeur réalisée et la valeur prévue, est distribuée selon une loi normale.

$$Y_{T+k} - Y_T^*(k) = e_T(k) \sim N(0, \sigma^2 \sum_{j=0}^{k-1} b_j^2)$$

Dans un modèle de prévision classique, les coefficients b_j figurant dans la formule de prévision sont estimés en minimisant la variance de l'erreur de prévision. L'intervalle de prévision a pour extrémités :

$$Y_T^*(k) \pm \Pi^{-1}(1-\alpha/2)\sqrt{\left[\hat{\sigma}^2 \sum_{j=0}^{k-1} \hat{b}_j^2\right]}$$

où $\Pi^{-1}(1-\alpha/2)$ est la fonction de répartition inverse de la loi normale au seuil α = 0,05. La largeur de cet intervalle est constante pour une valeur de k donnée quel que soit t et vaut

$$2\,\Pi^{-1}(1-\alpha/2)\sqrt{\left[\hat{\sigma}^2 \sum_{j=0}^{k-1} \hat{b}_j^2\right]}$$

Dans le modèle envisagé ici, les bornes de l'intervalle de prévision sont établies à partir de la variance conditionnelle et variables dans le temps. Par suite, pour un modèle dans lequel les erreurs sont engendrées par un processus ARCH, les bornes seront données par :

$$Y_T^*(k) \pm \Pi^{-1}(1-\alpha/2)\sqrt{\hat{h}_T^2 \sum_{j=0}^{k-1} \hat{b}_j^2}$$

La largeur de ce nouvel intervalle est fonction de h_T et varie donc à chaque période. De ce fait, même sous l'hypothèse de marche aléatoire où la prévision est réduite au plus simple (x_t est le meilleur prédicteur de x_{t+1}), l'intervalle de la prévision est plus satisfaisant, car il tient compte à chaque période des mouvements régnant sur le marché. En période de calme l'intervalle de prévision calculé en T pour l'horizon k est réduit. En revanche, au cours d'une période agitée l'intervalle de prévision s'élargit. Nous voyons donc que l'emploi des processus ARCH permet, en termes de prévision, une meilleure approche de la réalité qu'un processus bruit blanc.

Section 4. Processus GARCH.

Ce processus a été introduit par BOLLERSLEV (1985) et constitue une "**G**énéralisation" du processus **ARCH**.

§1. Présentation.

Soit le modèle de régression dont l'équation à la période t est $Y_t = x'_t b + \varepsilon_t$

Y_t : variable expliquée

x_t : vecteur des k variables explicatives

b : vecteur des k coefficients inconnus

ε_t : résidu

Si on appelle I_{t-1} l'information disponible en t-1, le résidu défini conditionnellement à cette information est :

$$\varepsilon_t / I_{t-1} \sim N(0, h_t^2)$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2 = \alpha_0 + \tilde{\alpha}(L)\varepsilon_t^2 + \tilde{\beta}(L) h_t^2$$

$\tilde{\alpha}(L)$: polynôme de retard d'ordre p

$\tilde{\beta}(L)$: polynôme de retard d'ordre q

Si q = 0, on est en présence d'un processus ARCH(p); si p = q = 0, $\{\varepsilon_t\}$ est un bruit blanc. Dans tous les autres cas, il s'agit d'un GARCH(p,q).

La variance conditionnelle inclut le carré des p erreurs précédentes et les q variances conditionnelles passées; elle évolue de période en période et intègre les erreurs les plus récentes de manière à constituer un mécanisme adaptatif.

Si le polynôme $[1-\tilde{\beta}(L)]$ est inversible (donc si $[1-\tilde{\beta}(L)] = 0$ admet toutes ses racines hors du cercle unitaire), il est possible de réécrire ce processus comme un ARCH d'ordre infini, et on aura.

$$h_t^2 = \alpha_0 + \tilde{\alpha}(L)\varepsilon_t^2 + \tilde{\beta}(L)h_t^2$$

$$\left[1-\tilde{\beta}(L)\right] h_t^2 = \alpha_0 + \tilde{\alpha}(L)\varepsilon_t^2$$

$$h_t^2 = \left[1-\tilde{\beta}(L)\right]^{-1}\alpha_0 + \left[1-\tilde{\beta}(L)\right]^{-1}\tilde{\alpha}(L)\varepsilon_t^2 = \left[1-\tilde{\beta}(L)\right]^{-1}\alpha_0 + D(L)\varepsilon_t^2$$

Les termes du polynômes D(L) sont de la forme :

$$\delta_i = \alpha_i + \sum_{j-1}^{n} \beta_j \delta_{i-j} \qquad 1 \leq i \leq q$$

$$= \sum_{j-1}^{n} \beta_j \delta_{i-j} \qquad i > q \qquad\qquad n = \min(p, i-1)$$

§2. Stationnarité.

L'écriture de GARCH(p,q) en termes de ARCH(∞) permet de déduire facilement les conditions de stationnarité d'un processus GARCH(p,q). En effet, pour un ordre P suffisamment grand tel que ARCH(P) soit une bonne approximation de ARCH(∞), le processus est stationnaire du second ordre si :

$$\sum_{i=1}^{p} \alpha_i < 1 \quad .$$

Cette condition devient D(1)<1, et comme $D(1) = \tilde{\alpha}(1)\,[1 - \tilde{\beta}(1)]^{-1} < 1$, la condition finale est la suivante :

$$\tilde{\alpha}(1) + \tilde{\beta}(1) < 1$$

§3. Moments non conditionnels d'un processus GARCH(1,1).

Pour un processus GARCH(1,1) on a

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2$$

$$\alpha_0 > 0 \quad \alpha_1 \geq 0 \quad \beta_1 \geq 0$$

Le processus est stationnaire si $\alpha_1 + \beta_1 < 1$

Pour ce processus, le 2m^{ème} moment existe si
$\mu(\alpha, \beta_1, m) = \sum_{j=0}^{m} C_m^j \, a_j \, \alpha_1^j \, \beta_1^{m-j} < 1$
avec $a_0 = 1$ et $a_j = \prod_{i=1}^{j} (2j-1)$

Les moments d'ordre 2m sont obtenus par la formule récursive suivante:

$$E(\varepsilon_t^{2m}) = a_m \left[\sum_{n=0}^{m-1} (1/a_n)\, E(\varepsilon_t^{2n})\, \alpha_0^{m-n}\, C_m^{m-n}\, \mu(\alpha, \beta_1, m) \right] . \left[1 - \mu(\alpha, \beta_1, m)\right]^{-1}$$

Si le 2m^{ème} moment existe, alors $E(\varepsilon_t^{2m-1}) = 0$.

Un des moments particuliers, la variance non conditionnelle, est donné par :

$$\sigma^2 = \frac{\alpha_0}{1 - \delta_1} = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

§4. Estimation d'un modèle de régression GARCH(p,q).

Préalablement à l'estimation, la procédure d'identification d'un GARCH(p,q) est la même que celle qui est proposée par BOX et JENKINS pour les modèles ARIMA.

L'estimation du modèle est alors faite en employant l'algorithme de BERNDT, HALL, HALL et HAUSMAN, basé sur le principe du maximum de vraisemblance.

A. Le problème.

Soit le modèle de régression :

$$Y_t = x'_t b + \varepsilon_t \qquad \varepsilon_t \sim (0,\sigma^2)$$

$$\varepsilon_t / I_{t-1} \sim N(0,h_t^2)$$

$$h_t^2 = \alpha_0 + \tilde{\alpha}(L)\varepsilon_t^2 + \tilde{\beta}(L)h_t^2$$

Le logarithme de la fonction de vraisemblance conditionnelle à l'instant t est, à une constante près :

$$l_t(\theta) = -\frac{1}{2}\log\left(h_t^2\right) - \frac{1}{2}\varepsilon_t^2\, h_t^{-2}$$

où $\theta = (\alpha_0, \alpha_1,\ldots, \alpha_p, \beta_1,\ldots, \beta_q, b')$ représente le vecteur des p+1+q+k paramètres inconnus. Pour l'ensemble de la période, la fonction est :

$$L(\theta) = \sum_{t=\max(p,q)}^{T} l_t(\theta)$$

Les conditions du premier ordre de la maximisation sont établies en posant $w' = (\alpha_0,\alpha_1,\ldots,\alpha_p,\beta_1,\ldots,\beta_q)$ et $\theta' = (w' \;\; b')$

$$\frac{\delta L}{\delta w} = \sum_{t=1}^{T} \frac{1}{2h_t^2} \frac{\delta h_t^2}{\delta w} \left[\frac{\varepsilon_t^2}{h_t^2} - 1\right] = 0$$

$$\frac{\delta L}{\delta b} = \sum_{t=1}^{T} \frac{\varepsilon_t}{h_t^2} X_t + \sum_{t=1}^{T} \frac{1}{2h_t^2} \frac{\delta h_t^2}{\delta b} \left[\frac{\varepsilon_t^2}{h_t^2} - 1 \right] = 0$$

Ces équations ne sont pas linéaires par rapport aux paramètres inconnus. Il faut donc trouver une méthode numérique de résolution.

B. Résolution par l'algorithme de BHHH.

L'algorithme de BERNDT, HALL, HALL et HAUSSMAN (1974) utilise la matrice d'information $R_{\theta\theta}$. L'apparition de termes récursifs dans les équations de dérivées secondes (ENGLE, 1982) complique la procédure et ne permet pas d'utiliser l'algorithme du score basé sur les dérivées premières de la log-vraisemblance. La procédure itérative de BHHH permet d'obtenir les estimateurs du maximum de vraisemblance malgré ce problème. Sous l'hypothèse que la matrice d'information est bloc-diagonale, w et b peuvent être estimés séparément sur la base d'une estimation de l'autre.

La procédure commence par l'estimation du vecteur b par les MCO; on en déduit les résidus estimés qui sont introduits dans les conditions du premier ordre et dans l'expression de la matrice d'information. A partir de là, la procédure itérative va évoluer en calculant alternativement w et b. En posant ζ successivement égal à w puis b on définit $\zeta^{(i)}$ l'estimation de ζ à l'itération i; à l'itération i+1 on aura :

$$\zeta^{(i+1)} = \zeta^{(i)} + \mu_i \left[R_{\zeta\zeta}^{(i)}\right] \sum_{t-1}^{T} \frac{\delta l_t^{(i)}}{\delta \zeta^{(i)}}$$

où $R_{\zeta\zeta}^{(i)}$ est une estimation faite à l'étape i du bloc en ζ de la matrice d'information; μ_i est un scalaire représentant le poids de l'itération, il permet à la procédure de se diriger vers le maximum de la log-vraisemblance en évitant les maxima locaux. μ_i s'obtient en régressant un vecteur d'ordre T dont toutes les composantes sont égales à 1 sur $\delta l_t/\delta\zeta$. WEISS (1982) a établi les propriétés asymptotiques de cet estimateur :

$$\sqrt{T}\left(\hat{\zeta} - \zeta\right) \sim N(0, R_{\zeta\zeta}^{-1})$$

La variance non conditionnelle des erreurs ε_t est constante et donnée par :

$$\sigma^2 = \frac{\alpha_0}{1 - D(1)} = \frac{\alpha_0}{1 - \tilde{\alpha}(1) - \tilde{\beta}(1)}$$

Elle est estimée à la suite de la dernière itération de l'algorithme précédent.

§5. Test de GARCH(p,q).

Dans le cas d'une hétéroscédasticité conditionnelle supposée, on va effectuer un test permettant de choisir entre un GARCH et un ARCH .Le test le plus approprié est celui du multiplicateur de Lagrange. La variance conditionnelle va être décomposée afin de tester l'hypothèse nulle d'une erreur ARCH(p) contre l'hypothèse alternative d'une erreur GARCH(p,q).

Considérons l'expression

$$h_t^2 = \alpha_0 + \tilde{\alpha}(L)\, \varepsilon_t^2 + \tilde{\beta}(L)\, h_t^2 = z_t' w$$

Si $\beta_j = 0$ pour j=1,..,q alors le processus est un ARCH(p). La réécriture de cette variance, en séparant les deux composantes, permet de poser facilement le test.

Soit

$$h_t^2 = z_t' w = z_{1t}' w_1 + z_{2t}' w_2$$

$$\text{où} \quad z_{1t}' = \left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-p}^2\right)$$

$$z_{2t}' = \left(h_{t-1}^2, \ldots, h_{t-q}^2\right)$$

$$w_1' = \left(\alpha_0, \alpha_1, \ldots, \alpha_p\right)$$

$$w_2' = \left(\beta_1, \ldots, \beta_q\right)$$

L'hypothèse nulle selon laquelle $\{\varepsilon_t\}$ est un processus ARCH(p), est que w_2 soit nul.

$H_0 : w_2 = 0$ (ARCH)

$H_1 : w_2$ différent de 0 (GARCH)

Sous H_0 la statistique du test du multiplicateur de Lagrange est :

$$L_{LM} = (f'_0 Z_0 (Z'_0 Z_0)^{-1} Z'_0 f_0)/2$$

avec

$$f_0' = \left[\frac{\varepsilon_1^2}{h_1^2} - 1, \ldots, \frac{\varepsilon_T^2}{h_T^2} - 1\right] \qquad \text{vecteur } (1,T)$$

$$Z'_0 = \left[h_1 \frac{\delta h_1^2}{\delta w}, \ldots, h_T \frac{d h_T^2}{\delta w} \right] \qquad \text{matrice } (p+1+q,T)$$

Sous H_0, L_{LM} suit un chi-deux à q degrés de liberté (q est le nombre d'éléments de w_2). Finalement $L_{LM} = T.R^2$, où R^2 est le coefficient de corrélation multiple obtenu lors de la première régression de $Y = Xb + \varepsilon$ par les MCO. On ne peut tester GARCH que contre l'hypothèse nulle de ARCH. En effet, sous l'hypothèse nulle de bruit blanc, $Z'_0 Z_0$ est singulière et la statistique L_{LM} ne peut pas être construite.

§6. Classification des processus GARCH.

Depuis 1985, différentes spécifications de GARCH ont été développées. DROST et NIJMAN les ont répertoriées en 1990 dans un article qu'ils ont présenté lors de la conférence sur les processus ARCH, organisée à Paris en Juin 1990.

A. GARCH fort.

Un processus $\{\varepsilon_t\}$ est généré par un GARCH(p,q) fort si α_0, $\tilde{\alpha}(L)$ et $\tilde{\beta}(L)$ sont tels que :

$$\frac{\varepsilon_t}{\sqrt{h_t^2}} \sim \text{iid}(0,1)$$

ENGLE (1982) et BOLLERSLEV (1985) adoptent cette définition avec une distribution normale. BOLLERSLEV (1987) la reprend avec une distribution de Student.

B. GARCH semi-fort.

Un processus $\{\varepsilon_t\}$ est généré par un GARCH(p,q) semi-fort si α_0, $\tilde{\alpha}(L)$ et $\tilde{\beta}(L)$ sont tels que :

$$E\left[\varepsilon_t / \varepsilon_{t-1}, \varepsilon_{t-2} \ldots\right] = 0$$

$$E\left[\varepsilon_t^2 / \varepsilon_{t-1}, \varepsilon_{t-2} \ldots\right] = h_t^2$$

Cette définition a été adoptée par WEISS (1984). Un GARCH fort sera un GARCH semi-fort. La réciproque n'est pas obligatoirement vraie.

C. GARCH faible.

Un processus $\{\varepsilon_t\}$ est généré par un GARCH(p,q) faible si α_0, $\tilde{\alpha}(L)$ et $\tilde{\beta}(L)$ sont tels que :

$$\mathrm{Pbb}\left[\varepsilon_t / \varepsilon_{t-1}, \varepsilon_{t-2} \ldots\right] = 0$$

$$\mathrm{Pbb}\left[\varepsilon_t^2 / \varepsilon_{t-1}, \varepsilon_{t-2} \ldots\right] = h_t^2$$

Un GARCH semi-fort est un GARCH faible, la réciproque n'est pas vraie.

Les processus ARCH et GARCH ont entraîné de multiples travaux et le développement d'autres processus à variance conditionnelle de forme autorégressive ; nous allons présenter les principaux.

Section 5. Autres processus.

Cette section est consacrée à une brève présentation des diverses extensions du modèle ARCH les plus couramment utilisées dans la littérature économique et financière.

§1. Processus de type ARCH in Mean

Soit ε_t l'erreur d'une équation de la forme $Y_t = x_t' b + \varepsilon_t$. Le processus $\{\varepsilon_t / I_{t-1}\}$ d'erreur définie conditionnellement à l'information passée est de la forme :

$$\varepsilon_t / I_{t-1} \sim N(0, h_t^2)$$

A. ARCH-M et GARCH-M.

ENGLE, LILIEN et ROBINS (1987) élargissent cette notion en considérant que le moment conditionnel de premier ordre est non nul et fonction du moment conditionnel du deuxième ordre.

Le processus sera un ARCH-M si :

$$Y_t = x_t' b + f(h_t^2) + \varepsilon_t \qquad (1)$$

où ε_t est un ARCH. Autrement dit le modèle est le suivant :

$$Y_t / I_{t-1} \sim N(\mu_t, h_t^2) \qquad \text{avec} \qquad \mu_t = x_t' b + f(h_t^2)$$
$$h_t^2 = \alpha_0 + \tilde{\alpha}(L)\varepsilon_t^2$$

$f(.)$ est une fonction quelconque. Il est toujours tenu compte de l'hétéroscédasticité conditionnelle mais en plus l'expression de l'espérance conditionnelle inclut la variance conditionnelle. Ce processus peut être très utilisé en finance où il paraît logique de supposer que les cours boursiers sont fonction de leur volatilité. Ce processus est nommé ARCH-M (*ARCH in Mean*). Le processus GARCH-M est alors défini de la manière suivante :

$\{Y_t\}$ est un processus GARCH-M obtenu en réécrivant la variance conditionnelle sous la forme :

$$h_t^2 = \alpha_0 + \tilde{\alpha}(L)\varepsilon_t^2 + \tilde{\beta}(L)h_t^2 \qquad (1bis)$$

B. GARCH-DM et GARCH-DLM

1. GARCH-DM.

COCCO et PARUOLO (1990) définissent le processus *GARCH-Difference in Mean* :

$$Y_t = x_t' b + \delta(h_t^2 - h_{t-1}^2) + \varepsilon_t \qquad (2)$$

où x_t est un ensemble de variables explicatives et h_t^2 la variance conditionnelle d'un processus GARCH classique. Le processus $\{Y_t\}$ est plus sensible aux différences de volatilité qu'à la volatilité elle-même comme cela était supposé dans le processus GARCH-M.

2. GARCH-DLM.

Le processus *GARCH-Distributed Lag in Mean* (GARCH-DLM) (COCCO, PARUOLO, 1990) est défini à partir d'un processus $\{Y_t\}$ représenté à la période t par l'équation suivante:

$$Y_t = x_t'b + \Delta(L)\left(h_t^2 - h_{t-1}^2\right) + \varepsilon_t \qquad (3)$$

$\Delta(L)$ est un polynôme scalaire de retard d'ordre r

$$\varepsilon_t / I_{t-1} \sim N(0, h_t^2)$$

Le processus GARCH-DM est un cas particulier de GARCH-DLM dans lequel $\Delta_0 = \delta$ et $\Delta_i = 0$ pour $i = 1, ..., r$.

3. Tests

Différents tests peuvent être effectués sur ces processus afin de permettre de différencier des séries quelquefois proches l'une de l'autre :

H_0	H_1	test à effectuer	équation
GARCH	GARCH-M	$\delta = 0$	(1bis)
GARCH	GARCH-DM	$\delta = 0$	(2)
GARCH	GARCH-DLM	$\Delta_i = 0 \ \forall \ i = 0, ..., r$	(3)
GARCH-M	GARCH-DLM	$\Delta_i = 0 \ \forall \ i = 1, ..., r$	(3)

Bien entendu les tests ARCH et GARCH contre l'hypothèse de bruit blanc sont les mêmes que ceux définis précédemment et relèvent du test du multiplicateur de Lagrange.

§2. Integrated GARCH (IGARCH).

Il peut arriver que le processus GARCH soit impossible à définir. C'est à ce problème qu'il est répondu ici, à partir d'un processus GARCH (1,1)

$$\varepsilon_t / I_{t-1} \sim N(0, h_t^2)$$

$$E\left[\varepsilon_t^2 / I_{t-1}\right] = h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2$$

Si $\alpha_1 + \beta_1 = 1$ la variance non conditionnelle σ^2 n'existe pas; on a affaire à un processus caractérisé par le fait que le dénominateur de σ^2 est nul. Dans cette situation, l'expression de la variance conditionnelle est la suivante :

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + (1-\alpha_1) h_{t-1}^2$$
$$= \alpha_0 + \alpha_1 \left(\varepsilon_{t-1}^2 - h_{t-1}^2 \right) + h_{t-1}^2$$

ENGLE et BOLLERSLEV (1986) appellent ce modèle *Integrated GARCH* (IGARCH). Pour un GARCH(p,q), cette situation apparaît lorsque

$$\sum_{i=1}^{p+q} (\alpha_i + \beta_i) = 1$$

§3 Treshold ARCH (TARCH).

Des travaux existent sur des processus AR dans lesquels les variables sont constantes par morceaux c'est à dire varient lorsque la grandeur atteint un seuil fixé à l'avance; il s'agit des processus *Treshold AR* (TAR). Ces travaux ont été menés notamment par TONG et LIM (1980), CHAN, PETRUCELLI et WOOLFORD (1985). ZAKOIAN (1990) s'inspire de ces études et utilise cette notion de modèles autorégressifs à seuils qu'il intègre à l'étude des processus ARCH de la manière suivante. Il définit le modèle TARCH en modifiant la forme de la variance conditionnelle :

$$h_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i^+ \varepsilon_{t-1}^+ + \sum_{i=1}^{p} \alpha_i^- \varepsilon_{t-1}^-$$

avec $\varepsilon_t^+ = \mathrm{Max}(\varepsilon_t, 0)$

$\varepsilon_t^- = \mathrm{Min}(\varepsilon_t, 0)$

En finance, un tel modèle permet de spécifier l'idée de dissymétrie de l'information ou, plus exactement, un comportement différent de la part des agents face à l'information selon que cette dernière reflète une hausse ou une baisse des cours boursiers. MONFORT et GOURIEROUX (1990) ont repris cette notion en l'appliquant à des données qualitatives (modèle QTARCH).

§4 Modèle EGARCH (Exponential GARCH).

Le modèle EGARCH (NELSON, 1991) introduit une deuxième source de non linéarité en remplaçant la variance conditionnelle par son logarithme. Le modèle dans lequel le terme aléatoire est un EGARCH est le suivant :

$$y_t = x_t' \beta + \varepsilon_t$$
$$\varepsilon_t \text{ EGARCH avec } \varepsilon_t = u_t h_t$$

L'erreur est alors définie par :

$$\varepsilon_t / I_{t-1} \sim N\left(0, h_t^2\right)$$

avec

$$\ln\left(h_t^2\right) = \alpha + \sum_{k=1}^{\infty} \beta_k \, g(u_k)$$

NELSON propose la formulation suivante pour g()

$$g(u_t) = \theta u_t + \gamma\left[\,|u_t| - E|u_t|\,\right]$$

Une expression générale pour un EGARCH(p,q) pourrait être la suivante :

$$\ln\left(h_t^2\right) = \alpha_0 + \alpha(L)\left[\theta u_t + \gamma\left[\,|u_t| - E|u_t|\,\right]\right] + \delta(L)\ln\left(h_t^2\right)$$

où α_1 est normalisé à 1 et $\alpha(L)$ et $\delta(L)$ sont deux polynômes scalaires de retards respectifs p et q.

Une forme particulière EGARCH(1,1) estimable est proposée par HSIEH (1991) et donnée par :

$$\ln\left(h_t^2\right) = \alpha_0 + \psi \frac{\varepsilon_{t-1}}{h_{t-1}} + \lambda \left|\frac{\varepsilon_{t-1}}{h_{t-1}}\right| + \delta \ln\left(h_{t-1}^2\right)$$

§5 Typologie des modèles de type ARCH.

BERA et LEE (1990) proposent une structure générale pour les processus à variance conditionnelle de forme autorégressive. Leurs travaux permettent d'avoir une vision générale de tous les processus développés depuis ARCH. Soit $\{\varepsilon_t\}$ un processus aléatoire et I_{t-1} l'information disponible en t-1; on définit le processus conditionnel :

$\varepsilon_t/I_{t-1} \sim N(\mu_t, h_t^2)$

Un modèle de régression peut être spécifié comme suit :

$\varepsilon_t = Y_t - x_t' \beta$

$$\varepsilon_t = \sum_{i=1}^{p} \left(\Phi_i + \eta_{it}\right) \varepsilon_{t-i} + \sum_{j=1}^{q} \alpha_{jt} h_{t-j} + u_t$$

β et $\Phi_.$ sont des paramètres fixes inconnus.

η_{it} et α_{jt} sont des coefficients aléatoires.

Ils posent alors les hypothèses suivantes:

1) $\eta_t = (\eta_{1t}, \ldots, \eta_{pt})'$ est un vecteur aléatoire d'ordre p. Les vecteurs η_t sont supposés indépendants et identiquement distribués avec :

$E(\eta_t) = 0.$

$E\left(\eta_t \eta_t'\right) = \Sigma_p$ où Σ_p est une matrice carrée d'ordre p semi-définie positive indépendante de t.

2) $\alpha_t = (\alpha_{1t}, \ldots, \alpha_{qt})'$ est un vecteur aléatoire d'ordre q. les vecteurs α_t sont indépendants et identiquement distribués avec :

$E(\alpha_t) = 0$

$E\left(\alpha_t \alpha_t'\right) = \Delta_c$ où Δ_c est une matrice carrée d'ordre q semi-définie positive, diagonale et indépendante de t, de terme générale δ_q.

3) u_t est un bruit blanc $(0,\sigma^2)$.

4) η_t, u_t et δ_t sont mutuellement indépendants.

Les moments conditionnels de premier et de deuxième ordre peuvent alors être définis comme suit:

$$E(\varepsilon_t/I_{t-1}) = \mu_t = \phi' \underline{\varepsilon}_t \qquad (1)$$

$$V(\varepsilon_t/I_t) = h_t^2 = \underline{\varepsilon}_t' \Sigma_p \underline{\varepsilon}_t + \underline{h}_t' \Delta_q \underline{h}_t \qquad (2)$$

avec $\phi = (\phi_1 \ldots \phi_p)'$ vecteur de p paramètres inconnus

$\underline{\varepsilon}_t = (\varepsilon_{t-1} \ldots \varepsilon_{t-p})'$ vecteur des p retards de ε_t

$\underline{h}_t = (h_{t-1} \ldots h_{t-q})'$ vecteur des q retards de l'écart-type conditionnel h_t

Les équations (1) et (2) définissent une forme générale d'autocorrélation et d'hétéroscédasticité conditionnelle.

Les différents processus sont alors caractérisés de la manière suivante :

* ARCH si
  $\phi = 0$ $\Delta_q = 0$ et Σ_p est diagonale
* GARCH (Generalised ARCH) si
  $\phi = 0$ Δ_q et Σ_p sont diagonales
* AARCH (Augmented ARCH) si
  $\phi = 0$ $\Delta_q = 0$ et Σ_p est quelconque
* GAARCH (Generalised Augmented ARCH) si
  $\phi = 0$ Δ_q diagonale et Σ_p quelconque
* AR si
  ϕ non nul $\Delta_q = 0$ $\Sigma_p = 0$

Les propriétés générales sont développées dans l'article de BERA et LEE (1990).

Conclusions

Ce document, volontairement technique, décrit un ensemble de processus ayant de nombreuses applications en particulier dans l'analyse des séries financières (ALEXANDRE, 1991). L'accent a été mis ici sur une tentative de survol de la littérature concernant les "descendants" du processus ARCH défini par ENGLE en 1982. Nous avons d'abord montré l'intérêt d'un tel processus, ensuite nous avons abordé une généralisation de ce modèle. Dans les deux cas les procédures d'estimation ont été développées. La dernière section, consacrée aux dérivés du modèle ARCH, prouve l'impact que celui-ci a eu sur la littérature économétrique des années 80. Il est ainsi mis en évidence que la modélisation en terme d'hétéroscédasticité conditionnelle est utilisée par un nombre de plus en plus grand de chercheurs et qu'il était temps de faire le point.

Décembre 1992

BIBLIOGRAPHIE

ALEXANDRE H., (1991), La Quasi Marche Aléatoire, CREGO, vol. 9103, 25 pages.

BELSLEY D., (1979), On the efficient computation of non linear full-information Maximum Likelihood estimator, Papier présenté au 'European Meetings of the Econometric Society' à Athènes.

BERA A.K. LEE S., (1990), On the formulation of a general structure for conditional heteroskedasticity, papier présent à la conférence sur les processus ARCH Paris, INSEE, Juin, 1990, 39 pages.

BERNDT E.K., HALL B.,. HALL R.E. HAUSMAN J.A., (1974), Estimation inference in nonlinear structural models, Annals of economic and Social Measure, vol 4, pages 653-665.

BOLLERSLEV T, (1986), Generalised AutoRegressive Conditional Heteroskedasticity, Journals of Econometrics, vol 31, pages 307-327.

BOLLERSLEV T., ENGLE R.F., (1986), Modelling the persistence of conditional variances, Econometric Review, vol. 5, pages 1-50.

BOX G. JENKINS J.M., (1976), Time series analysis: forecasting and control, Holden-Day San-Francisco CA.

BREUSCH T.S., PAGAN A.R., (1978), A simple test for heteroskedasticity and random coefficient variation, Econometrica, vol. 46, pages 1287-1294.

CHANG K.S., PETRUCELLI F., TONG H., WOOLFORD, (1985), A multiple treshold AR(1) model, Journal of Applied Probability, vol 2, pages 267-279.

COCCO F. PARUOLO P., (1990), Volatility persistence and the Italian risk premium: parametric and non-parametric evaluation, papier présent à la conférence sur les processus ARCH Paris, INSEE, Juin 1990, 30 pages.

COX D.R. HINKLEY D.V., (1974), Theoritical Statistics, Londres, Chapman and hall.

CROWDER M.J., (1976), Maximum likelihood estimation for dependent observations, Journal of the Royal Statistical Society, vol. 38, pages 45-53.

DROST F.C., NIJMAN T.E., (1990), Temporal aggregation of GARCH processes, papier présenté à la conférence sur les processus ARCH, Paris, INSEE, Juin 1990, 15 pages.

ENGLE R.F., (1979), A general approach to the construction of model diagnostic based upon the Lagrange Multiplier principle, Université de Califorie, San Diego.

ENGLE R.F., (1982), Autoregressive Conditionnal Heteroskedasticity with Estimates of the Variance of U.K. inflation, Econometrica, vol. 50, pages 987-1008.

ENGLE R.F., LILIEN D. M. ROBINS R.P.,(1987), Estimating time varying premia in the term structure: the ARCH-M model, Econometrica, vol. 55, pages 391-407.

FAMA E.F., (1970), Efficient capital markets: a review of theory and empirical work, Journal of Finance, vol. 25, pages 383-417.

FAMA E.F., (1965), The behaviour of stock market prices, Journal of Business , vol. 38, pages 34-105.

GODFREY L.G., (1978), Testing again general autoregressive and moving-average error model when the regressor include lagged dependant variables, Econometrica, vol. 46, pages 1293-1302.

GOURIEROUX C., MONFORT A., (1990 a), Séries temporelles et modèles dynamiques, Economica (Collection ESA), 780 pages.

GOURIEROUX C., MONFORT A., (1990 b), Qualitative treshold ARCH models, Documents de travail du CREST, n° 9009, 43 pages.

GRANGER C.W., ANDERSEN A., (1978), An introduction to bilinear time-series models, Göttingen: Vandenhoeck and Ruprecht.

GRANGER C.W., NEWBOLD P., (1977), Forecasting economic time-series models, Academic Press, New-York.

HSIEH D.A., (1991), Chaos and nonlinear dynamics : Application to financial markets, The Journal of Finance, vol 46, pages 1839-1877.

MAC NEES S.S., (1979), The forecasting record for the 1970's, New-England Economic Review, Sept-Oct 79, pages 33-53.

MANDELBROT B., (1963), The variation of certain speculative prices, Journal of Business, vol. 36, pages 394-419.

MUSSE M., (1979), Empirical regularities in the behavior of exchanges rates and theories of the foreign exchange market, Carnegie-Rochester conference series on public policy, vol. 11, pages 9-57.

NELSON D.B., (1991), Conditional heteroskedasticity in asset returns : a new approach, Econometrica, vol. 59, n° 2, pages 347-370.

TONG M., LIM K.S., (1980), Treshold autoregression, limit cycles and cyclical data, Journal of Royal Statistical Society, vol. 42, pages 245-292.

WEISS A.A., (1984), ARMA models with ARCH errors, Journal of Time Series Analysis, vol. 5, pages 125-143.

ZAKOIAN J.M., (1990), ARCH models with Treshold, papier présenté à la conférence sur les processus ARCH, Paris, INSEE, Juin 1990, 43 pages.