

Un test d'hétéroscédasticité conditionnelle inspiré de la modélisation en termes de réseaux neuronaux artificiels

Renaud CAULET, Anne PEGUIN-FEISSOLLE *

RÉSUMÉ. – Ce papier considère un test d'hétéroscédasticité conditionnelle basé sur la méthode des réseaux neuronaux artificiels et en compare les performances avec des tests standards, à l'aide de simulations de Monte-Carlo. L'hypothèse alternative d'hétéroscédasticité conditionnelle est représentée par une variance conditionnelle de forme neuronale ; le test du Multiplicateur de Lagrange qui en découle permet de détecter une grande variété de formes d'hétéroscédasticité conditionnelle. Les résultats des simulations, présentés sous forme graphique, montrent que ce test est relativement performant.

A Test for Conditional Heteroscedasticity Based on Artificial Neuronal Networks

ABSTRACT. – This paper considers a test for conditional heteroscedasticity based on artificial neural networks and compares its performance with some standard tests using a Monte Carlo study. The conditionally heteroscedastic alternative hypothesis is represented by a conditional variance with a neural specification; the Lagrange Multiplier test can detect a large class of conditional heteroscedasticity. The results of the simulation experiments, that are entirely presented in graphical form, show that its relative performance is encouraging.

* R. CAULET : GREQAM ; A. PEGUIN-FEISSOLLE : GREQAM-CNRS

Nous remercions F. APRAHAMIAN, M. KAMSTRA, J. MACKINNON, T. TERÄSVIRTA et un rapporteur anonyme pour les remarques qu'ils nous ont apportées sur des versions antérieures de ce papier. Bien entendu, les erreurs ou imperfections qui pourraient subsister sont de notre seule responsabilité.

1 Introduction

La modélisation économétrique en termes d'hétéroscédasticité conditionnelle s'est largement répandue dans la littérature depuis l'article d'ENGLE en 1982 introduisant les modèles *ARCH* (*AutoRegressive Conditionally Heteroscedastic*). De nombreuses variantes en ont été proposées et appliquées, particulièrement dans le domaine financier (*cf.* BOLLERSLEV, CHOU et KRONER [1992], BERA et HIGGINS [1993] et BOLLERSLEV, ENGLE et NELSON [1994]). Se sont développés parallèlement des tests d'hétéroscédasticité conditionnelle, par exemple le test du Multiplicateur de Lagrange proposé par ENGLE [1982], ou des variantes des tests de Box-Pierce ou de Ljung-Box (*cf.* MCLEOD et LI [1989]).

Nous nous intéressons ici à un test d'hétéroscédasticité conditionnelle, inspiré des techniques de modélisation en termes de réseaux neuronaux artificiels, ces dernières ayant la propriété d'approcher relativement bien une large classe de fonctions. Sur cette base, l'hypothèse alternative d'hétéroscédasticité conditionnelle est représentée par une variance conditionnelle de forme neuronale ; le test du Multiplicateur de Lagrange qui est construit sur ce modèle permet de détecter relativement bien une grande variété de formes d'hétéroscédasticité conditionnelle. Récemment, LEE, WHITE et GRANGER [1993] ont proposé un test de linéarité basé sur ces techniques.

Nous comparons, par simulation et sur un large éventail de spécifications possibles de la variance conditionnelle, les performances de ce test avec celles de trois autres tests d'hétéroscédasticité conditionnelle, dont l'un est un test neuronal élaboré par KAMSTRA [1993], mais utilisant une statistique différente.

Le plan de l'article est le suivant. Nous présentons, d'abord, le test d'hétéroscédasticité conditionnelle, ensuite, les résultats des simulations et, enfin, une conclusion fait l'objet de la dernière partie.

2 Présentation du test neuronal

Le modèle de régression avec erreurs conditionnellement hétéroscédastiques s'écrit :

$$(1) \quad y_t = x_t' \alpha + \epsilon_t \quad t = 1, \dots, T$$

avec

$$(2) \quad \epsilon_t | x_t, J_{t-1} \sim IIN(0, h_t).$$

J_t représente l'ensemble d'information en t , comprenant les variables endogènes retardées et les variables exogènes ; y_t est la variable endogène et x_t est

un vecteur $k \times 1$ d'éléments de J_t ; α est un vecteur $k \times 1$ de paramètres inconnus. La variance conditionnelle s'exprime généralement comme

$$(3) \quad h_t = h(J_{t-1}, \delta)$$

où δ est un vecteur de paramètres inconnus ; par exemple, nous avons un modèle $ARCH(q)$ si $h_t = \delta_0 + \delta_1 \epsilon_{t-1}^2 + \dots + \delta_q \epsilon_{t-q}^2$ ou un modèle $GARCH(p, q)$ si

$$h_t = \xi_0 + \sum_{j=1}^q \xi_j \epsilon_{t-j}^2 + \sum_{j=1}^p \omega_j h_{t-j}$$

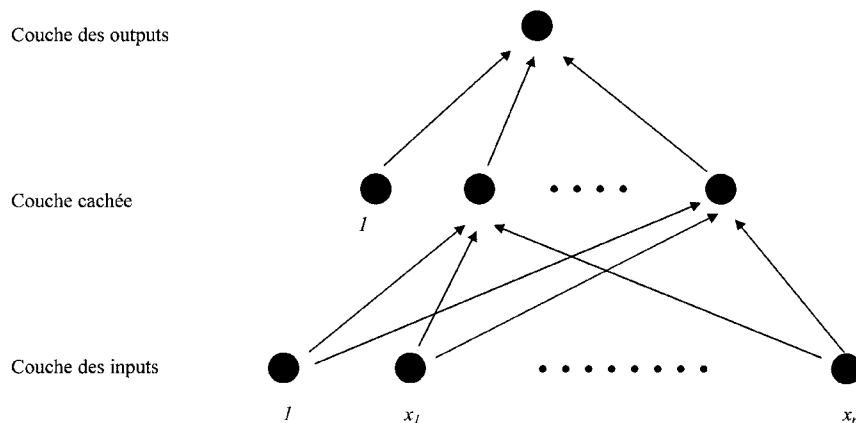
(δ est alors le vecteur constitué des paramètres $\xi_0, \xi_1, \dots, \xi_q, \omega_1, \dots, \omega_p$).

L'hypothèse nulle à laquelle nous nous intéressons est l'hypothèse d'homoscedasticité, à savoir celle caractérisée par la constance de la variance conditionnelle h_t . Nous allons considérer un test particulier d'hétéroscédasticité conditionnelle dans lequel l'hypothèse alternative est caractérisée par une variance conditionnelle h_t de forme neuronale.

2.1 Bref aperçu des réseaux neuronaux artificiels

Les modèles de neurones artificiels ont été introduits par les cognitivistes afin de représenter la façon dont l'information est traitée dans le cerveau (*cf.*, par exemple, STERN [1996]). L'architecture du réseau que nous choisissons de présenter et d'utiliser ensuite pour la construction du test est celle d'un réseau très simple, avec une seule couche cachée ; elle est décrite dans la figure 1. Ce réseau se compose de trois couches : la couche d'inputs, la couche cachée et la couche d'outputs. Les n unités d'input envoient des signaux notés x_1, \dots, x_n , amplifiés ou atténués par des facteurs de pondération γ_{ij} , à q unités ou noeuds cachés ; les unités cachées somment les signaux et génèrent une fonction d'activation ψ . Les signaux des unités cachées, pondérés par des poids β_j pour $j = 1, \dots, q$, sont envoyés ensuite dans la couche d'output ; ici, un seul

FIGURE 1
Réseau de neurones à n inputs, une seule couche cachée et un seul output



output est considéré. En faisant intervenir à chaque couche la constante 1, cet *output* unique s'exprime de la façon suivante, en posant $\beta = (\beta_0, \beta_1, \dots, \beta_q)'$:

$$(4) \quad f(w_t, \beta, \gamma_1, \dots, \gamma_q) = \beta_0 + \sum_{j=1}^q \beta_j \psi(w_t' \gamma_j)$$

où $w_t = (1, x_1, \dots, x_n)'$ est le vecteur d'*inputs*, $\gamma_j = (\gamma_{j0}, \gamma_{j1}, \dots, \gamma_{jn})'$ est le vecteur des poids de la couche d'*inputs*, pour $j = 1, \dots, q$, et ψ est une fonction non linéaire donnée de \mathbb{R} vers \mathbb{R} ; très souvent, ψ est la fonction logistique de la forme : $\psi(\lambda) = (1 + e^{-\lambda})^{-1}$, $\lambda \in \mathbb{R}$. D'autres fonctions pourraient être utilisées, comme la fonction tangente hyperbolique.

D'autres réseaux plus complexes ont été étudiés, par exemple, avec plusieurs couches cachées, des liens directs entre les deux couches extrêmes ou des récurrences entre diverses couches.

La littérature sur les réseaux neuronaux artificiels a montré que des fonctions de la forme (4) fournissent des approximations relativement bonnes de fonctions arbitraires de w_t , pour q suffisamment grand et un choix convenable des vecteurs β et γ_j , $j = 1, \dots, q$ (voir, par exemple, HORNIK, STINCHCOMBE et WHITE [1989 et 1990], STINCHCOMBE et WHITE [1989], CYBENKO [1989], CAROLL et DICKINSON [1989], parmi d'autres). C'est en raison de cette flexibilité que les techniques des réseaux neuronaux artificiels ont attiré l'attention de certains auteurs (voir, KUAN et WHITE [1994]). Cette qualité particulière de tels réseaux, qui ont d'ailleurs été appelés approximateurs universels, va être exploitée dans ce qui suit pour construire un test permettant de détecter des formes diverses d'hétéroscédasticité conditionnelle.

2.2 Construction du test neuronal

Nous allons considérer le modèle de régression (1) et (2) dans lequel la variance conditionnelle h_t est une fonction neuronale de la forme suivante :

$$(5) \quad h_t = \beta_0 + \sum_{j=1}^q \beta_j (w_t' \gamma_j),$$

où $w_t = (1, \epsilon_{t-1}, \dots, \epsilon_{t-n})'$ est le vecteur des *inputs* ; les poids des différentes couches sont représentés comme ci-dessus par $\gamma_{j0}, \gamma_{j1}, \dots$ et γ_{jn} pour $j = 1, \dots, q$, ainsi que par β_0, β_1, \dots , et β_q ; les γ_j se définissent par $\gamma_j = (\gamma_{j0}, \gamma_{j1}, \dots, \gamma_{jn})$ pour tout j et ψ est la fonction logistique. Les retards du terme d'erreur jouent le rôle des variables d'*input* du réseau. Ainsi, la variance conditionnelle h_t s'écrit

$$(6) \quad h_t = \beta_0 + \sum_{j=1}^q \frac{\beta_j}{1 + e^{-(\gamma_{j0} + \gamma_{j1}\epsilon_{t-1} + \dots + \gamma_{jn}\epsilon_{t-n})}}.$$

Il est intéressant de noter, ici, que cette formulation permet aux chocs d'avoir des effets asymétriques : les effets des chocs positifs et négatifs ne seront pas les mêmes, contrairement à certaines formulations traditionnelles

de modèles à hétéroscédasticité conditionnelle telles que les formulations ARCH et GARCH ; divers modèles (comme les modèles EGARCH et TARARCH, par exemple) ont d'ailleurs été proposés dans la littérature afin de capturer ce type d'effets. Nous reviendrons sur ceci lorsque nous commenterons certains exemples de simulation.

Certains auteurs comme DONALDSON et KAMSTRA [1997] ont estimé des modèles ARCH utilisant des fonctions neuronales dans lesquels la variance conditionnelle a une forme se rapprochant de (6) ; notre objectif est différent puisque nous introduisons une variance conditionnelle de forme neuronale afin de construire un test qui permette de détecter des classes assez larges d'hétéroscédasticité, comme nous le verrons dans les expériences de simulation.

Le test d'hétéroscédasticité conditionnelle que nous exposons ici s'intéresse à l'hypothèse nulle suivante :

$$(7) \quad H_0 : \beta_j = 0 \quad j = 1, \dots, q$$

dans le modèle donné par (1), (2) et (6), pour un choix particulier des vecteurs $\gamma_1, \dots, \gamma_q$ et pour un nombre q d'unités cachées. Suivant LEE, WHITE et GRANGER [1993], les poids γ_{ji} de la couche d'inputs sont choisis *a priori*, indépendamment des erreurs passées, pour $q \in \mathbb{N}$; ce choix permet de résoudre le problème de la non-identification des γ_{ji} sous l'hypothèse nulle.

En supposant l'indépendance et la normalité conditionnelle des erreurs, la vraisemblance logarithmique s'écrit

$$(8) \quad \ln L(\theta) = \sum_{t=1}^T l_t(y_t | \theta)$$

avec

$$(9) \quad l_t(y_t | \theta) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln h_t - \frac{1}{2h_t} (y_t - m_t)^2$$

où $\theta' = (\alpha', \beta') = (\alpha', \beta_0, \beta^{*'})$ avec $\beta^* = (\beta_1, \dots, \beta_q)'$ et $\beta' = (\beta_0, \beta^{*'})$; par ailleurs, $m_t = x_t' \alpha$ et h_t est donné par (6). Il faut noter, ici, que la vraisemblance logarithmique définie en (8) et (9) est également conditionnelle à certaines valeurs initiales des variables x_t et y_t et qu'en outre, elle dépend également des x_t , pour $t = 1, \dots, T$. Nous supposons, également, satisfaites les conditions assurant les propriétés asymptotiques du maximum de vraisemblance.

Nous utilisons la procédure de test du Multiplicateur de Lagrange : cette procédure nécessite seulement l'estimation du modèle sous l'hypothèse nulle.

Rappelons que les γ_{ji} sont supposés connus *a priori*. La statistique LM qui sert à tester

$$(10) \quad H_0 : \beta^* = 0 \text{ contre } H_1 : \beta^* \neq 0,$$

est donnée par (voir Annexe) :

$$(11) \quad NNLM = \frac{1}{2} \left(\sum_t \left(\frac{\widehat{\epsilon}_t^2}{\widehat{\sigma}^2} - 1 \right) \frac{\partial h_t}{\partial \beta^{*'}} \right)'$$

$$\left(\sum_t \frac{\partial h_t}{\partial \beta^*} \frac{\partial h_t}{\partial \beta^{*'}} - \frac{1}{T} \left(\sum_t \frac{\partial h_t}{\partial \beta^*} \right) \left(\sum_t \frac{\partial h_t}{\partial \beta^{*'}} \right) \right)^{-1}$$

$$\left(\sum_t \left(\frac{\widehat{\epsilon}_t^2}{\widehat{\sigma}^2} - 1 \right) \frac{\partial h_t}{\partial \beta^{*'}} \right);$$

cette expression est évaluée en $\widehat{\theta}_R$, l'estimateur restreint de θ , calculé sous H_0 ; $\widehat{\epsilon}_t$, $t = 1, \dots, T$, est le résidu estimé du modèle (1) sous H_0 :

$$\widehat{\epsilon}_t = y_t - x_t' \widehat{\alpha} \quad t = 1, \dots, T$$

et $\widehat{\sigma}^2 = \sum_t \widehat{\epsilon}_t^2 / T$. La statistique $NNLM$ est asymptotiquement distribuée comme une χ_q^2 sous H_0 .

Considérons la régression auxiliaire suivante :

$$(12) \quad \frac{\widehat{\epsilon}_t^2}{\widehat{\sigma}^2} - 1 = \beta_0 + \frac{\partial h_t}{\partial \beta^{*'}} \beta^* + \eta_t \quad t = 1, \dots, T,$$

où $\frac{\partial h_t}{\partial \beta^*}$ est le vecteur $q \times 1$

$$(13) \quad \frac{\partial h_t}{\partial \beta_j^*} = \frac{1}{1 + e^{-(\gamma_{j0} + \gamma_{j1} \widehat{\epsilon}_{t-1} + \dots + \gamma_{jn} \widehat{\epsilon}_{t-n})}} \quad j = 1, \dots, q$$

qui est donc évalué en $\widehat{\alpha}$, l'estimateur de α obtenu sous H_0 .

Posons $\epsilon^* = \left(\frac{\widehat{\epsilon}_1^2}{\widehat{\sigma}^2} - 1, \dots, \frac{\widehat{\epsilon}_T^2}{\widehat{\sigma}^2} - 1 \right)'$ et $\eta = (\eta_1, \dots, \eta_T)'$, deux vecteurs

$T \times 1$, et $W^* = (i, X^*)$, une matrice $T \times (q + 1)$, où i est le vecteur $T \times 1$ tel que $i = (1, \dots, 1)'$ et X^* est la matrice $T \times q$ définie par

$$X^* = \begin{pmatrix} \frac{\partial h_t}{\partial \beta_j^{*'}} \\ \vdots \end{pmatrix}_{\substack{t=1, \dots, T \\ j=1, \dots, q}}.$$

La régression (12) s'exprime encore

$$(14) \quad \epsilon^* = W^* \delta + \eta.$$

Les différents éléments de (11) peuvent s'écrire

$$\sum_t \left(\frac{\widehat{\epsilon}_t^2}{\widehat{\sigma}^2} - 1 \right) \frac{\partial h_t}{\partial \beta^{*'}} = X^{*'} \epsilon^*$$

et

$$\sum_t \frac{\partial h_t}{\partial \beta^*} \frac{\partial h_t}{\partial \beta^{*'}} - \frac{1}{T} \left(\sum_t \frac{\partial h_t}{\partial \beta^*} \right) \left(\sum_t \frac{\partial h_t}{\partial \beta^{*'}} \right) = X^{*'} M_i X^*$$

où $M_i = I - \frac{1}{T} ii'$. En utilisant les propriétés des matrices partitionnées et le

fait que $i' \epsilon^* = \sum \frac{\widehat{\epsilon}_t^2}{\widehat{\sigma}^2} - T = 0$, il est facile de montrer que

$$\begin{aligned} NNLM &= \frac{1}{2} \epsilon^{*'} X^* (X^{*'} M_i X^*)^{-1} X^{*'} \epsilon^* \\ &= \frac{1}{2} \epsilon^{*'} W^* (W^{*'} W^*)^{-1} W^{*'} \epsilon^*. \end{aligned}$$

Ainsi, nous pouvons écrire $NNLM$ comme la moitié de la somme expliquée des carrés de la régression auxiliaire représentée par (12) ou (14), c'est-à-dire :

$$(15) \quad NNLM = \frac{1}{2} \widehat{\epsilon}^{*'} \widehat{\epsilon}^*$$

où $\widehat{\epsilon}^*$ est le vecteur $T \times 1$ des variables endogènes estimées dans la régression auxiliaire (12).

3 Les expériences de Monte-Carlo

Des séries univariées seront engendrées à partir de différents modèles choisis pour représenter un éventail assez large de formes d'hétéroscédasticité conditionnelle. Les méthodes de Monte-Carlo sont utilisées pour comparer le test neuronal donné par (15) avec d'autres tests.

3.1 Tests alternatifs

Pour étudier les propriétés de taille et de puissance du test neuronal en petit échantillon, nous les comparons avec celles de trois autres tests :

– La statistique de test LM contre des erreurs $ARCH(n)$ est donnée par (voir formule (35) dans ENGLE [1982]) :

$$(16) \quad ALM = \frac{1}{2} \epsilon^{*'} Z (Z' Z)^{-1} Z' \epsilon^*$$

où $\epsilon^* = \left(\frac{\hat{\epsilon}_1^2}{\hat{\sigma}^2} - 1, \dots, \frac{\hat{\epsilon}_T^2}{\hat{\sigma}^2} - 1 \right)'$, $Z' = (z_1, \dots, z_T)$, et $z'_t = (1, \hat{\epsilon}_{t-1}^2, \dots, \hat{\epsilon}_{t-n}^2)$.

– Une statistique asymptotiquement équivalente (provenant du fait que $\text{plim } \epsilon^{*'} \epsilon^* / T = 2$ sous l'hypothèse de normalité) est la suivante (voir formule (36) dans ENGLE [1982]):

$$(17) \quad AR^2 = T \frac{\epsilon^{*'} Z (Z' Z)^{-1} Z' \epsilon^*}{\epsilon^{*'} \epsilon^*} = TR^2$$

où R^2 est le carré du coefficient de corrélation multiple entre ϵ^* et Z .

– Le test de Kamstra (KAMSTRA [1993]) est un test neuronal d'hétéroscédasticité conditionnelle où la variance conditionnelle prend la forme :

$$h_t^* = \beta_0 + \sum_{j=1}^q \frac{\beta_j}{1 + e^{-(\gamma_{j0} + \gamma_{j1} \epsilon_{t-1}^2 + \dots + \gamma_{jn} \epsilon_{t-n}^2)}};$$

la statistique qui est utilisée est la suivante :

$$(18) \quad NNR^2 = TR^2$$

où R^2 est le carré du coefficient de corrélation multiple de la régression auxiliaire de $\hat{\epsilon}_t^2$ sur une constante et le vecteur $\frac{\partial h_t^*}{\partial \beta^*}$ évalué sous H_0 (KAMSTRA [1993] recommande de prendre les plus grandes composantes principales de la matrice des régresseurs dans cette régression auxiliaire).

Les trois statistiques de test données par (16), (17) et (18) suivent asymptotiquement une distribution du χ^2 avec, respectivement, n (pour (16), (17)) et q (pour (18)) degrés de liberté sous l'hypothèse nulle.

Les notations des différentes statistiques de test ($NNLM$, ALM , AR^2 and NNR^2) illustrent bien l'idée qu'il existe deux alternatives (A représentant l'alternative $ARCH$ et NN représentant l'alternative neuronale) et deux formes de test (la forme LM et la forme TR^2).

3.2 Mise en place des simulations

Les expériences de simulation nous permettent de comparer les performances de ces différents tests. Nous présentons les résultats concernant la taille et la puissance en utilisant la présentation graphique de DAVIDSON et MACKINNON [1993 et 1994]. Elle conduit à des graphes qui sont beaucoup

plus faciles à interpréter que les tableaux habituellement utilisés pour présenter ce type de résultats. Ces graphes sont construits à partir de la fonction de distribution empirique des p -values¹ de réalisations simulées τ_j , $j = 1, \dots, N$, d'une statistique de test τ . Appelons p_j la p -value associée à τ_j ; il s'agit de la probabilité d'observer une valeur de τ plus grande que τ_j relativement à une distribution $F(\tau)$. La fonction de distribution empirique des p_j est définie par

$$(19) \quad \widehat{F}(x_i) = \frac{1}{N} \sum_{j=1}^N I(p_j \leq x_i)$$

où I est la fonction indicatrice suivante :

$$(20) \quad I(p_j \leq x_i) = \begin{cases} 1 & \text{si } p_j \leq x_i \\ 0 & \text{sinon} \end{cases}$$

et x_i est un point de l'intervalle $[0, 1]$. Puisque N est généralement très grand, DAVIDSON et MACKINNON [1994, p. 2] conseillent d'évaluer $\widehat{F}(x_i)$, donné par (19), en m points x_i , $i = 1, \dots, m$, qui devront fournir une représentation raisonnable de l'intervalle $[0, 1]$; étant donné qu'il est difficile d'établir de façon catégorique la valeur de m et le choix des x_i , ces auteurs recommandent de prendre

$$(21) \quad x_i = 0.002, 0.004, \dots, 0.01, 0.02, \dots, 0.99, 0.992, \dots, 0.998 \quad (m = 107)$$

qui, d'après de nombreuses expériences, s'avère être un ensemble tout à fait parcimonieux.

En ce qui concerne la taille des différents tests, nous savons que si la distribution de τ est correcte, chacun des p_j doit être distribué uniformément sur l'intervalle $[0, 1]$; par conséquent, le graphe des $\widehat{F}(x_i)$ contre x_i doit être proche de la droite à 45° , ou encore le graphe des $\widehat{F}(x_i) - x_i$ contre x_i doit être proche de l'axe horizontal.

En ce qui concerne la puissance des différents tests, c'est-à-dire lorsque les données sont engendrées sous l'hypothèse alternative, nous donnerons les courbes taille-puissance; elles représentent le graphe de la puissance contre la vraie taille de chaque test: l'axe horizontal représente \widehat{F} , la fonction de distribution empirique lorsque les données sont engendrées sous l'hypothèse nulle, et l'axe vertical représente \widehat{F}^* , la fonction de distribution empirique lorsque les données sont engendrées sous l'hypothèse alternative, c'est-à-dire selon un modèle conditionnellement hétéroscédastique. Ainsi, les courbes taille-puissance sont engendrées selon une taille correcte (DAVIDSON et MACKINNON [1994, p. 11]).

En outre, pour ce qui concerne le test neuronal, les poids γ_{ji} des unités cachées sont engendrés aléatoirement, pour chaque itération, selon une distribution uniforme sur $[-2, 2]$ et les variables y_t et x_t sont normées sur $[0, 1]$,

1. Nous préférons utiliser, ici, le terme anglais de p -value plutôt que sa traduction française.

FIGURE 2

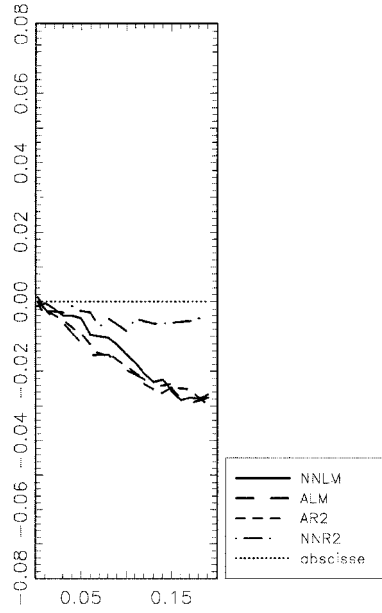


FIGURE 3

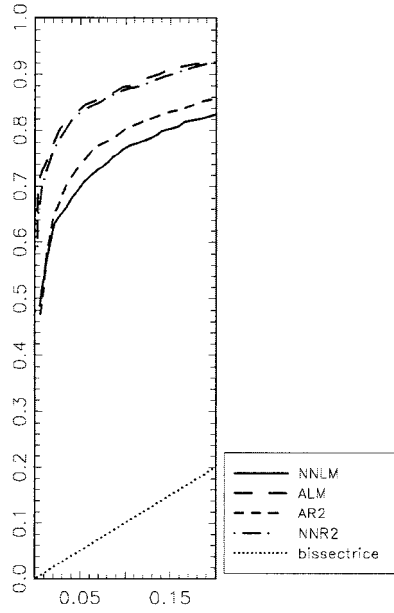


FIGURE 4

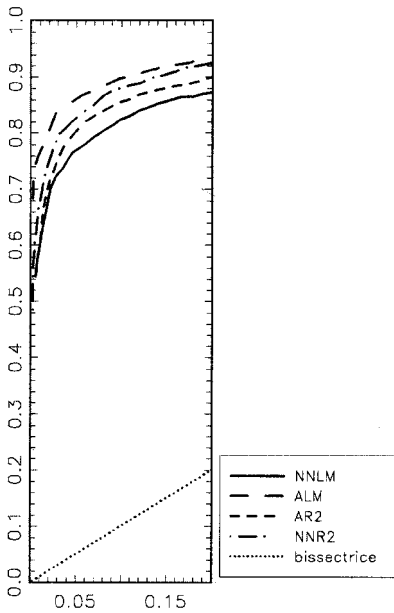
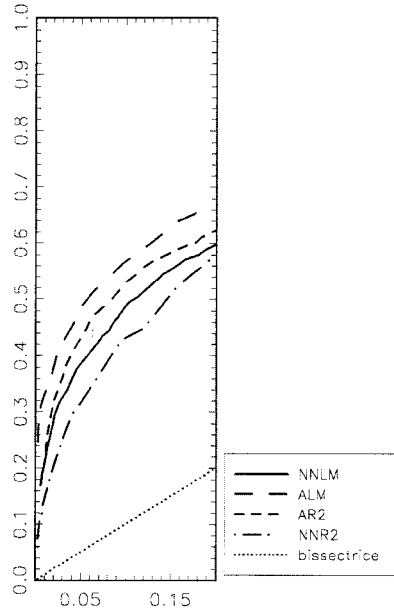


FIGURE 5



suivant LEE, WHITE et GRANGER [1993]. Par ailleurs, les statistiques $NNLM$ et NNR^2 sont calculées en utilisant les $q^* = 4$ plus grandes composantes principales au lieu des variables originelles dans les régressions auxiliaires respectives, ce qui permet d'éviter les problèmes posés parfois par leur nombre trop important ou leur trop forte corrélation ; le nombre choisi d'unités cachées est $q = 10$.

3.3 Résultats des simulations sous l'hypothèse nulle

Les données sont engendrées à partir du modèle normal homoscédastique suivant :

$$(22) \quad y_t = 0.25 + 0.5x_t + \epsilon_t \quad t = 1, \dots, T$$

où $\epsilon_t \sim IIN(0,1)$ et $x_t = 0.7x_{t-1} + v_t$ avec $v_t \sim IIN(0,1)$. Le nombre d'itérations, N , est égal à 2 500 ; le nombre d'inputs est $n = 3$ (plus la constante).

Nous ne donnons pas les graphes de la fonction de distribution empirique $\widehat{F}(x_i)$ contre x_i car ils sont tous pratiquement confondus avec la droite à 45° . La figure 2 montre les graphes de $\widehat{F}(x_i) - x_i$ contre x_i pour $T = 200$. Bien que ces écarts $\widehat{F}(x_i) - x_i$ soient faibles, il apparaît, néanmoins, que tous les tests sur-rejettent systématiquement l'hypothèse nulle pour des tailles très faibles, qui sont celles qui nous intéressent généralement.

3.4 Résultats des simulations sous l'hypothèse alternative

Afin de comparer la puissance des tests d'hétéroscédasticité conditionnelle, nous traçons les courbes de taille-puissance, c'est-à-dire le lieu des points $(\widehat{F}(x_i), \widehat{F}^*(x_i))$ où les valeurs des x_i sont données par (21), $\widehat{F}(x_i)$ étant la fonction de distribution empirique engendrée par le modèle linéaire (22) et $\widehat{F}^*(x_i)$ étant celle engendrée par un processus appartenant à l'hypothèse alternative. Suivant DAVIDSON et MACKINNON [1994], nous réduisons les erreurs expérimentales en utilisant le même ensemble de nombres aléatoires dans toutes les expériences. Par ailleurs, les échantillons artificiels sont construits avec plus d'observations qu'il n'est nécessaire, la dernière partie seulement étant effectivement utilisée.

3.4.1 Comparaison des différents tests dans divers cas d'hétéroscédasticité conditionnelle

L'hypothèse alternative est ainsi représentée par différents modèles, choisis pour représenter un certain éventail de situations conditionnellement hétéroscédastiques. Le nombre d'itérations est $N = 2500$ et le nombre d'observations

FIGURE 6

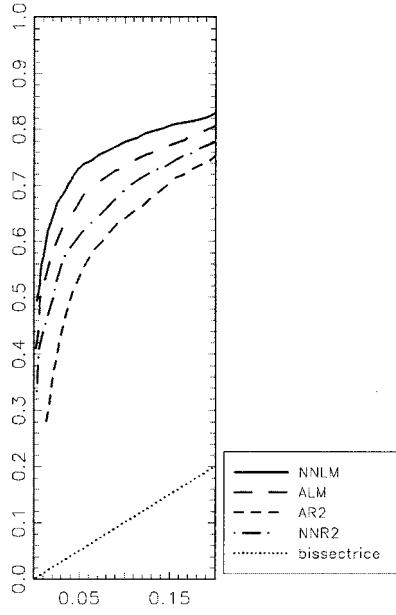


FIGURE 7

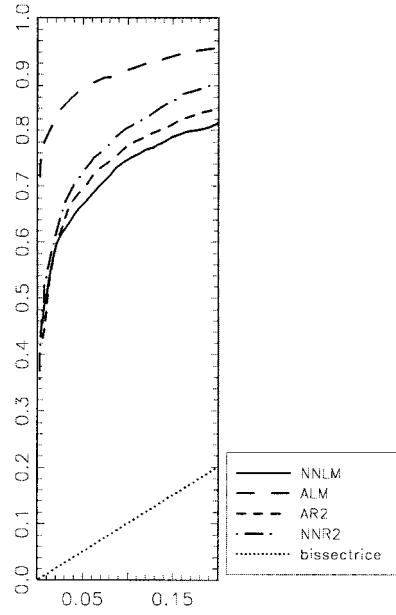


FIGURE 8

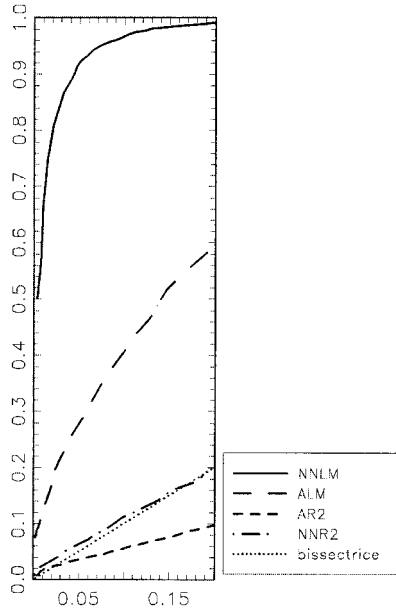
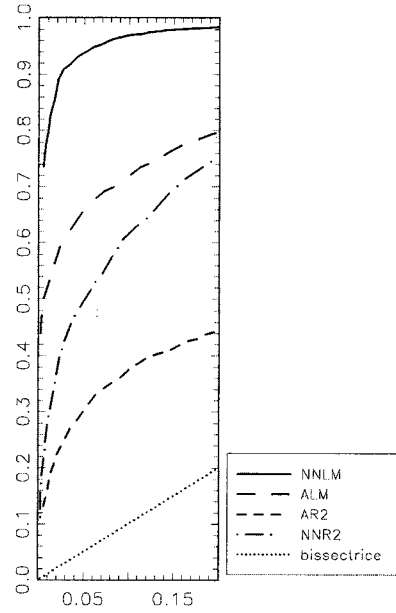


FIGURE 9



est $T = 100$. Le modèle de régression est donné par (22) avec différentes spécifications de h_t . Les erreurs ϵ_t sont générées sous la forme $\epsilon_t = v_t \sqrt{h_t}$ avec $v_t \sim IIN(0,1)$, $t = 1, \dots, T$.

Les différentes courbes taille-puissance correspondent à :

– cas 1 ($n = 3$) : Figure 3

$$(23) \quad h_t = 0.2 + 0.7\epsilon_{t-1}^2$$

– cas 2 ($n = 3$) : Figure 4

$$(24) \quad h_t = 0.2 + 0.3\epsilon_{t-1}^2 + 0.3\epsilon_{t-2}^2 + 0.3\epsilon_{t-3}^2$$

– cas 3 ($n = 3$) : Figure 5

$$(25) \quad h_t = 0.2 + 0.2\epsilon_{t-1}^2 + 0.7h_{t-1}$$

– cas 4 ($n = 1$) : Figure 6

$$(26) \quad h_t = \left[0.2 + 0.3 \left(\epsilon_{t-1}^2 \right)^{0.5} + 0.2 \left(\epsilon_{t-2}^2 \right)^{0.5} + 0.3 \left(\epsilon_{t-3}^2 \right)^{0.5} \right]^{0.5}$$

– cas 5 ($n = 3$) : Figure 7

$$(27) \quad h_t = \left[0.2 + 0.5 \left(\epsilon_{t-3}^2 \right)^2 + 0.5 \left(h_{t-3} \right)^2 \right]^{\frac{1}{2}}$$

– cas 6 ($n = 2$) : Figure 8

$$(28) \quad h_t = 10 + \frac{70}{1 + e^{0.2+0.4\epsilon_{t-1}+0.9\epsilon_{t-2}}} + \frac{90}{1 + e^{0.9+0.1\epsilon_{t-1}+0.5\epsilon_{t-2}}}$$

– cas 7 ($n = 3$) : Figure 9

$$(29) \quad h_t = \begin{cases} 0.01 + 5 |\epsilon_{t-3}| & \text{si } \epsilon_{t-1} \geq 0 \\ 1 / (1 + 2.5\epsilon_{t-3}^2) & \text{si } \epsilon_{t-1} < 0 \end{cases}$$

– cas 8 ($n = 3$) : Figure 10

$$(30) \quad h_t = \frac{1}{0.01 + 2.5\epsilon_{t-5}^2}$$

– cas 9 ($n = 2$) : Figure 11

$$(31) \quad h_t = e^{0.1+0.5\epsilon_{t-2}}$$

– cas 10 ($n = 3$) : Figure 12

FIGURE 10

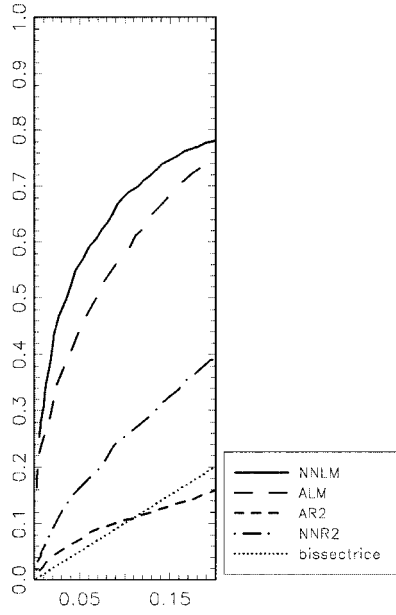


FIGURE 11

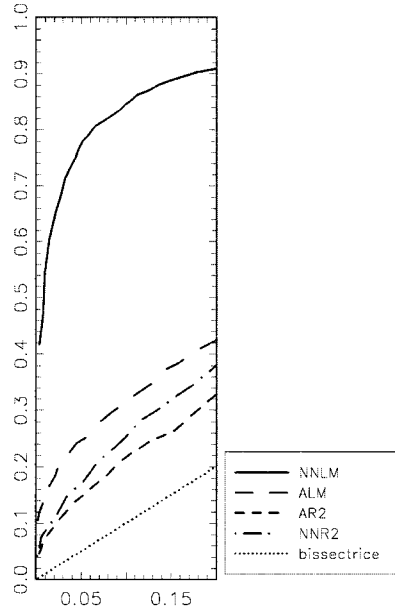


FIGURE 12

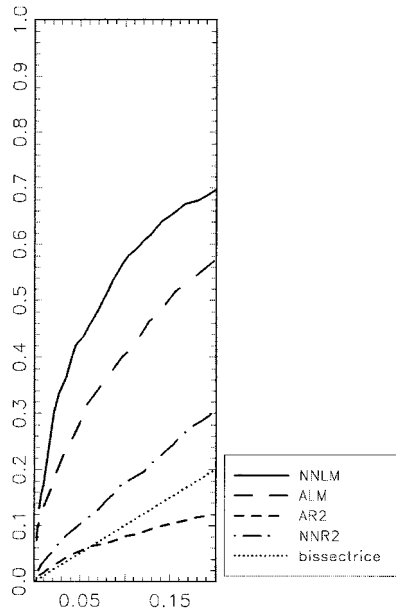
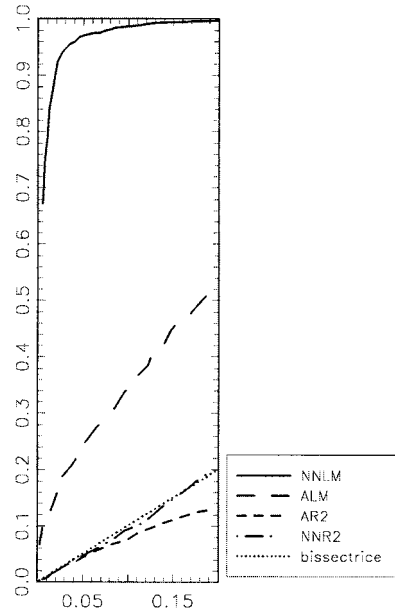


FIGURE 13



$$(32) \quad h_t = 1 + \epsilon_{t-5}^2 \left[1 - e^{-10\epsilon_{t-5}^2} \right]$$

– cas 11 ($n = 3$) :

$$(33) \quad h_t = \begin{cases} 10 & \text{si } \epsilon_{t-1} \geq 0 \\ 0.1 & \text{si } \epsilon_{t-1} < 0 \end{cases}$$

– cas 12 ($n = 3$) :

$$(34) \quad h_t = \begin{cases} 2 + 5 |\epsilon_{t-3}| & \text{si } \epsilon_{t-1} \geq 0 \\ 0.001 & \text{si } \epsilon_{t-1} < 0 \end{cases}$$

Les figures 3 à 14 illustrent ainsi le comportement des tests dans des cas différents. Certains correspondent à des formes fonctionnelles proposées dans la littérature : *ARCH* dans les cas 1 et 2, le cas 3 est un *GARCH*(1,1), le cas 4 est un *ARCH* non linéaire, le cas 5 un *GARCH* non linéaire. Les autres cas correspondent à des formes non standard d'hétéroscédasticité ; par exemple, les cas 7, 11 et 12 sont des formes à seuil et le cas 6 a une forme de réseau neuronal artificiel.

Par ailleurs, chaque expérience de simulation est menée avec $q^* = 4$, c'est-à-dire les statistiques *NNLM* et *NNR*² sont calculées avec 4 composantes principales dans la régression auxiliaire (plus la constante) sur la base de $q = 10$ unités sur la couche cachée. Pour chaque test, le nombre de retards des erreurs dans les régressions auxiliaires est $n = 3$ (hormis dans les cas 4 et 6). Ainsi, certains cas permettent de voir l'influence d'une mauvaise spécification de la forme fonctionnelle de l'hétéroscédasticité conditionnelle sur les performances des différents tests.

En effet, que n soit supérieur au nombre de retards de ϵ_t qui a servi à générer le modèle sous l'hypothèse alternative (comme dans les cas 1, 3 et 11), ou qu'il en soit inférieur (cas 4, 8 ou 10 par exemple), ceci ne nuit en rien aux performances comparées du test neuronal par rapport aux autres tests, puisque la statistique *NNLM* reste dans la plupart des cas au-dessus des autres statistiques ; ceci est important à noter puisque, dans les applications à des données réelles, on ne connaîtra pas généralement le vrai modèle ayant généré les données.

Les figures montrent dans leur ensemble que le test neuronal (représenté par la statistique *NNLM*) se comporte tout à fait bien ; il domine souvent les autres procédures de test, surtout lorsqu'il s'agit de formes d'hétéroscédasticité autres que les formes *ARCH* ou *GARCH* standard. En effet, *NNLM* n'est pas la meilleure statistique dans les figures 3, 4, 5 et 7 ; en revanche, pour tous les autres cas, la courbe correspondant à la statistique *NNLM* est nettement au-dessus des autres courbes.

Les figures 10, 12 et 13 montrent des cas particuliers où les tests, hormis le test neuronal et le test d'Engle donné par *ALM*, présentent de sérieux problèmes. En effet, les deux tests correspondant aux statistiques *AR*² et *NNR*² se comportent assez mal : leurs courbes taille-puissance sont très proches de la première bissectrice, ou bien leur taille est supérieure à leur puissance ; ils sont, ainsi, parfois biaisés, selon la façon dont les données sont engendrées.

FIGURE 14

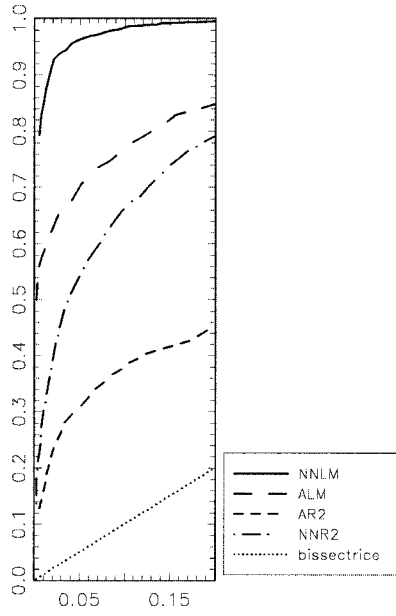


FIGURE 15

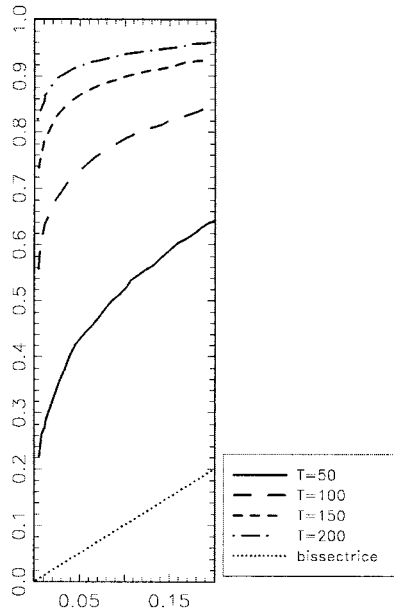


FIGURE 16

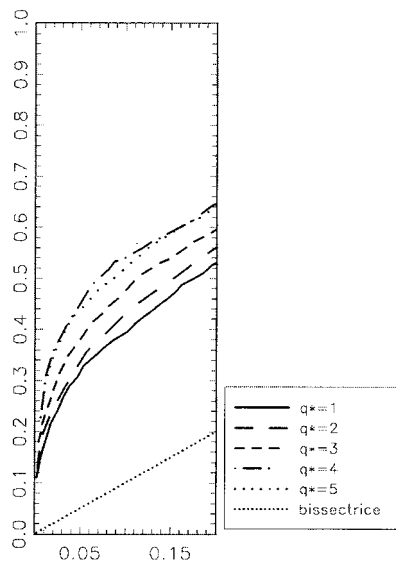
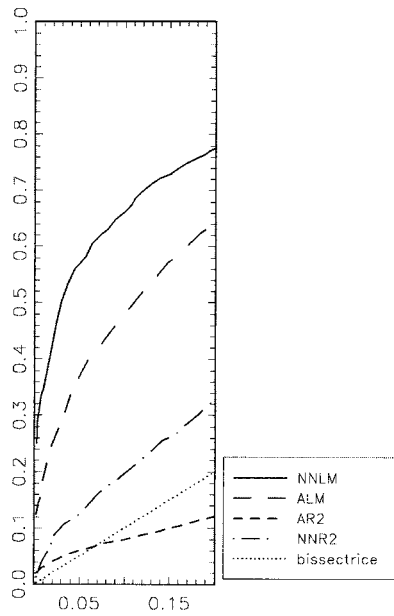


FIGURE 17



Enfin, nous avons noté dans le paragraphe 2.2 que la variance conditionnelle h_t , donnée par la formule (6) et ayant servi à construire la statistique $NNLM$, permettrait de tenir compte d'effets asymétriques possibles des ε_t selon leurs signes ; ceci se vérifie dans les cas 6, 7, 9, 11 et 12 qui incorporent de tels effets et pour lesquels la statistique $NNLM$ se détache, en effet nettement des autres statistiques.

3.4.2 Influence du nombre d'observations et du nombre de composantes principales sur le test neuronal

Les deux figures suivantes (15 et 16) exposent les courbes taille-puissance pour le test neuronal en fonction du nombre d'observations T et du nombre de composantes principales q^* . Pour ces deux figures, le modèle alternatif est donné par (22) et

$$(35) \quad h_t = \left[0.2 + 0.3 (\varepsilon_{t-1}^2)^{1.5} + 0.2 (\varepsilon_{t-2}^2)^{1.5} + 0.3 (\varepsilon_{t-3}^2)^{1.5} \right] \frac{1}{1.5} .$$

La figure 15, où $n = 3$ et $q^* = 3$, montre que, lorsque T augmente, les performances du test neuronal s'améliorent tout à fait. La figure 16, générée avec $n = 3$ et $T = 50$, montre que, dans cet exemple en tout cas, quatre composantes principales semblent donner les meilleures performances au test neuronal, mais ce nombre est dépendant des cas choisis. Dans les cas exposés ci-dessus, nous avons pris $q^* = 4$; LEE, WHITE et GRANGER [1993] utilisent deux ou trois composantes principales pour le test de linéarité qu'ils construisent.

3.4.3 Robustesse des différents tests à l'introduction d'une dynamique dans l'équation de régression

D'autres expériences de simulation ont été menées en introduisant la variable endogène retardée dans le modèle de régression ; nous avons ainsi étudié un modèle normal homoscédastique de la forme :

$$(36) \quad y_t = 0.25 + 0.5y_{t-1} + \varepsilon_t \quad t = 1, \dots, T$$

où $\varepsilon_t \sim IIN(0, h_t)$. Le nombre d'itérations, N , est toujours égal à 2 500. Différentes simulations menées avec diverses spécifications pour h_t ont donné des résultats semblables à ceux exposés ci-dessus, en termes de hiérarchie des différents tests. La figure 17 en montre un exemple où h_t correspond au cas 10 (équation (32)).

4 Conclusion

Nous venons d'exposer un test d'hétéroscédasticité conditionnelle inspiré de la modélisation reposant sur les réseaux neuronaux artificiels. Il est bien connu que les résultats des simulations de Monte-Carlo dépendent de nombreux facteurs. Néanmoins, le test neuronal (représenté par la statistique *NNLM*) montre dans ces simulations des performances supérieures à celles de la forme LM du test d'Engle, surtout lorsqu'on est en présence d'une forme d'hétéroscédasticité autre qu'un *ARCH* ou un *GARCH* standard ; il semble ainsi efficace dans un large éventail de cas possibles d'hétéroscédasticité conditionnelle, comme on peut le voir dans les différents graphes. On peut, donc, espérer qu'il permette de la détecter dans des applications économiques.

• Références bibliographiques

- BERA A.K., HIGGINS M.L. [1993]. – « ARCH Models: Properties, Estimation and Testing », *Journal of Economic Surveys*, 7, 4, pp. 305-366.
- BOLLERSLEV T., CHOU R.Y., KRONER K.F. [1992]. – « ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence », *Journal of Econometrics*, 52, 1/2, pp. 5-60.
- BOLLERSLEV T., ENGLE R.F., NELSON, D.B. [1994]. – « Arch Models », *The Handbook of Econometrics*, vol. 4, North-Holland.
- CAROLL S.M., DICKINSON B.W. [1989]. – « Construction of Neural Nets using the Radon Transform », in *Proceedings of the international joint conference on neural networks*, Washington, D.C., IEEE Press, I: pp. 607-611.
- CYBENKO G. [1989]. – « Approximation by Superpositions of a Sigmoid Function », *Mathematics of Control, Signals and Systems*, 2, pp. 303-314.
- DAVIDSON R., MACKINNON J.G. [1993]. – *Estimation and Inference in Econometrics*, Oxford University Press.
- DAVIDSON R., MACKINNON J.G. [1994]. – « Graphical Methods for Investigating the Size and Power of Hypothesis Tests », *GREQAM Working Paper* No 94A23, Marseille, France.
- DONALDSON R.G., KAMSTRA M. [1997]. – « An Artificial Neural Network-GARCH Model for International Stock Return Volatility », *Journal of Empirical Finance*, 4, pp. 17-46.
- ENGLE R.F. [1982]. – « Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United-Kingdom Inflation », *Econometrica*, 50, 4, pp. 987-1007.
- HORNIK K., STINCHCOMBE M., WHITE H. [1989]. – « Multi-layer Feedforward Networks are Universal Approximators », *Neural Networks*, 2, pp. 359-366.
- HORNIK K., STINCHCOMBE M., WHITE H. [1990]. – « Universal Approximation of an Unknown Mapping and its Derivatives Using Multi-layer Feedforward Networks », *NEURAL NETWORKS*, 3, pp. 551-560.
- HSIEH D.A. [1989]. – « Modeling Heteroscedasticity in Daily Foreign-Exchange Rates », *Journal of Business & Economic Statistics*, 7, 3, pp. 307-317.
- KAMSTRA M. [1993]. – « A Neural Network Test for Heteroskedasticity », *Working Paper*, Simon Fraser University, Burnaby, Canada.
- KUAN C.M., WHITE H. [1994]. – « Artificial Neural Networks: an Econometric Perspective », *Econometric Reviews*, 13, 1, pp. 1-91.
- LEE T.H., WHITE H., GRANGER C.W.J. [1993]. – « Testing for Neglected Nonlinearity in Time Series Models – A Comparison of Neural Network Methods and Alternative Tests », *Journal of Econometrics*, 56, pp. 269-290.
- MCLEOD A.I., LI W.K. [1989]. – « Diagnostic Checking ARMA Time Series Models using Squared-Residual Autocorrelations », *Journal of Time Series Analysis*, 4, 4, pp. 269-273.
- STERN H.S. [1996]. – « Neural Networks in Applied Statistics », *Technometrics*, 38, 3, pp. 205-214.
- STINCHCOMBE M., WHITE H. [1989]. – « Universal Approximation using Feedforward Networks with Non-sigmoid Hidden Layer Activation Functions », in *Proceedings of the international joint conference on neural networks*, Washington, D.C., IEEE Press, I: pp. 613-618.

ANNEXE

Sous l'hypothèse d'indépendance et de normalité conditionnelle des erreurs, la vraisemblance logarithmique conditionnelle $\ln L(\theta)$ est donnée par (8) et (9) et h_t est donnée par (6). La statistique LM pour tester l'hypothèse nulle représentée par (10) s'écrit

$$NNLM = \frac{1}{T} \left(\frac{\partial \ln L(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_R} \right)' (I(\hat{\theta}_R))^{-1} \left(\frac{\partial \ln L(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_R} \right)$$

où $\hat{\theta}_R = (\hat{\alpha}, \hat{\beta}_0, 0)$ est l'estimateur de θ évalué sous H_0 . $I(\theta)$ est la matrice d'information définie par

$$I(\theta) = -E \left[\frac{\partial^2 l_t(y_t | \theta)}{\partial \theta \partial \theta'} \right]$$

et peut s'écrire

$$I(\theta) = E \left[\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \theta} \frac{\partial h_t}{\partial \theta'} + \frac{1}{h_t} \frac{\partial m_t}{\partial \theta} \frac{\partial m_t}{\partial \theta'} \right]$$

ou

$$I(\theta) = \begin{bmatrix} I^{11}(\theta) & I^{12}(\theta) \\ I^{21}(\theta) & I^{22}(\theta) \end{bmatrix}$$

avec

$$I^{11}(\theta) = \begin{bmatrix} E \left[\frac{1}{h_t} \frac{\partial m_t}{\partial \alpha} \frac{\partial m_t}{\partial \alpha'} \right] & 0 \\ 0 & E \left[\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \beta_0} \frac{\partial h_t}{\partial \beta_0'} \right] \end{bmatrix},$$

$$I^{12}(\theta) = I^{21}(\theta)' = \begin{bmatrix} 0 \\ E \left[\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \beta_0} \frac{\partial h_t}{\partial \beta^{*'}} \right] \end{bmatrix}$$

et

$$I^{22}(\theta) = E \left[\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \beta^*} \frac{\partial h_t}{\partial \beta^{*'}} \right].$$

On peut montrer que

$$NNLM = \frac{1}{T} \left(\frac{\partial \ln L(\theta)}{\partial \beta^{*'}} \right)' \left(I^{22}(\theta) - I^{21}(\theta) (I^{11}(\theta))^{-1} I^{12}(\theta) \right)^{-1} \left(\frac{\partial \ln L(\theta)}{\partial \beta^{*'}} \right),$$

les différentes composantes étant évaluées en $\hat{\theta}_R$; les espérances dans les matrices I^{ij} seront remplacées par les moments empiriques pour obtenir la statistique suivante, que nous appelons également $NNLM$ afin de ne pas alourdir les notations :

$$NNLM = \frac{1}{2} \left(\sum_t \left(\frac{\hat{\epsilon}_t^2}{\hat{\sigma}^2} - 1 \right) \frac{\partial h_t}{\partial \beta^{*'}} \right)' \left(\sum_t \frac{\partial h_t}{\partial \beta^*} \frac{\partial h_t}{\partial \beta^{*'}} - \frac{1}{T} \left(\sum_t \frac{\partial h_t}{\partial \beta^*} \right) \left(\sum_t \frac{\partial h_t}{\partial \beta^{*'}} \right) \right)^{-1} \left(\sum_t \left(\frac{\hat{\epsilon}_t^2}{\hat{\sigma}^2} - 1 \right) \frac{\partial h_t}{\partial \beta^{*'}} \right).$$

