Journal of Economic Dynamics and Control 18 (1994) 931–955. North-Holland

Threshold heteroskedastic models

Jean-Michel Zakoian*
Université de Lille and CREST, Paris, France

First version July 1990, final version December 1992

In this paper we consider a modification of the classical ARCH models introduced by Engle (1982). In this modified model the conditional standard deviation is a piecewise linear function of past values of the white noise. This specific form allows different reactions of the volatility to different signs of the lagged errors. Stationarity conditions are derived. Maximum likelihood and least squares estimation are also considered. Finally an empirical example relating to the French CAC stock index is presented and several specifications are compared.

Key words: GARCH models; Asymmetries in volatility; Stationarity

JEL classification: C32; C51

1. Introduction

Time-varying volatility models have been developed for a long time to describe series of prices [e.g., Taylor (1986)]. Historically, heteroskedasticity used to be specified as a function of exogenous observable variables. However, such formulations proved too restrictive and it was necessary to introduce an endogenous dynamic in the specification of the variabilities.

This progression led to ARCH models (Autoregressive Conditionally Heteroskedastic) introduced by Engle (1982) and to their extensions: GARCH [Bollerslev (1986)] or ARCH-M models [Engle, Lilien, and Robins (1985)]. These formulations explicitly model time-varying conditional variances by relating them to variables known from the previous periods. In its classical formulation, the GARCH regression model is obtained by assuming an autoregressive-moving average equation on an observable variable Y, the conditional variance

Correspondence to: Jean-Michel Zakoian, CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France.

*I should like to thank Christian Gouriéroux and two anonymous referees for their comments and Roger Rabemananjara for the empirical example.

being expressed as a linear function of past squared innovations and of its past values,

$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 + \beta_1\sigma_{t-1}^2 + \cdots + \beta_p\sigma_{t-p}^2, \tag{1}$$

where $\alpha_0, \alpha_1, \ldots, \beta_1, \ldots, \beta_p$ are nonnegative. It proved to be an effective tool in modelling temporal behaviour of many economic variables such as inflation [Engle (1982)] or macroeconomic time series [Weiss (1984)]. In the financial context, ARCH models provide a good description of two important characteristics of the series: time-varying volatility and leptokurticity. From a technical point of view, the quadratic specification used for the conditional variance has many convenient properties. First of all, the parameter constraints implied by the variance positivity are quite simple. In addition, it provides an ARMA representation for the ε^2 process and therefore a very tractable model.

However, this specification of the conditional variance as a function of past squared innovations has an important limit. The very characteristic time paths of ARCH processes are well known, with periods of high volatility (corresponding to high past values of the noise, of any sign) and periods of low volatility. Hence only a modulus effect of the past values of ε has been taken into account in the quadratic specification. In other words, whether a shock at date $t-k$ is positive or negative, it has the same effect on the present volatility. However a certain asymmetry in the correlation between the present volatility and the past values of the series has been pointed out by several authors as another characteristic property of stock returns series. Christie (1982), Nelson (1991), and Schwert (1989) have noted that when the asset price is rising, the volatility tends to lower and conversely.

An interesting modelization of volatility is the Quadratic GARCH, proposed by Sentana (1990), which allows for any positive definite quadratic form of the past innovations. This model, which encompasses the GARCH, provides asymmetries in volatility. However, it seems less tractable than the classical formulations since the positivity constraints may be difficult to check and since the ARMA model in ε^2 is replaced by a nonlinear one.

To account for asymmetries in the conditional variance more directly, Nelson (1991) proposed the following specification:

$$\ln\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q}\beta_j \ln\sigma_{t-j}^2 + \sum_{k=1}^{\infty}\alpha_k[\theta Z_{t-k} + \gamma(|Z_{t-k}| - E|Z_{t-k}|)], \tag{2}$$

$$\varepsilon_t = \sigma_t Z_t,$$

where (Z_t) is an i.i.d. process. The main advantage of this formulation over the standard GARCH is the fact that it allows positive and negative shocks of equal size to have different impacts on volatility. Moreover, the multiplicative modelling of volatility does not imply positivity constraints on the parameters. An important difference with the GARCH is that the innovations are divided by their conditional standard deviations [using ε instead of Z in (2) would certainly make the model intractable]. As a consequence, persistence in volatility only depends on the β_j parameters. Therefore, the effect on the volatility of a shock ε on the observable variable seems more difficult to interpret.

The aim of this paper is to propose a new specification of volatility, keeping the GARCH tractability while allowing for asymmetry. To take the latter into account we use a more flexible lag structure in the volatility, based on the positive and negative parts of the innovation process. This process, named threshold ARCH, is described in section 2 and some of its main properties (linearity, stationarity) are developed. Section 3 is devoted to inference: maximum likelihood and least squares methods are proposed. In section 4 we make a comparison of several ARCH and GARCH specifications on the French CAC stock index. Section 5 concludes. The proofs are given in the appendices.

2. Threshold GARCH

A natural device for introducing asymmetries in volatility is to make it a function of the positive and negative parts of the innovation process. In a recent paper, Glosten, Jagannathan, and Runkle (1989) have estimated an extension of GARCH, in which the conditional variance is written as a linear function of the squared positive and negative parts of the noise. We adopt a somewhat different approach, based on a paper from Davidian and Carroll (1987) about variance function estimation. One of the most interesting results that they obtain (though variance is allowed to depend on directly observable variables) is that in the case of nonnormal distributions, absolute residuals yield more efficient variance estimates than squared residuals. Therefore we do not square the positive and negative parts of the noise, and we specify the conditional standard deviation instead of the conditional variance. To formalize it, let ε_t denote a real-valued discrete-time process, $\underline{\varepsilon_{t-1}} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ the information set (σ-field) of all information through time t, $\varepsilon_t^+ = \max(\varepsilon_t, 0)$ and $\varepsilon_t^- = \min(\varepsilon_t, 0)$ the positive and negative parts of ε_t. The Threshold GARCH (p, q) process is then given by

$$\varepsilon_t = \sigma_t Z_t, \tag{3}$$

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^+ \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^- + \sum_{j=1}^{p} \beta_j \sigma_{t-j}, \tag{4}$$

$$(Z_t) \text{ i.i.d.}, \quad EZ_t = 0, \quad VZ_t = 1, \quad Z_t \text{ independent of } \underline{\varepsilon_{t-1}} \text{ for all } t, \tag{5}$$

where $(\alpha_i^+)_{i=1,q}$, $(\alpha_i^-)_{i=1,q}$, and $(\beta_j)_{j=1,p}$ are real scalar sequences.

Our approach is closely related to the one developed by Tong (1990) concerning the conditional mean modelling. The conditional standard deviation in (4) is driven by linear combinations of past ε and σ variables, the regime at date t depending on the positions of the past innovations on the real axis.

Another advantage of modelling the scale variable σ_t rather than the conditional variance comes from the fact that no positivity constraints are necessary in the definition of the different variables: the conditional variance σ_t^2 is nonnegative by construction; hence a great simplification in the numerical inference procedures. However, as far as the probabilistic properties are concerned, the study is much more complicated when σ_t is not assumed to be positive. It follows that in this paper, we complete the model with the positivity constraints

$$\alpha_0 > 0, \quad \alpha_i^+ \geq 0, \quad \alpha_i^- \geq 0, \quad \beta_i \geq 0 \quad \text{for all } i.^1 \tag{6}$$

Among the threshold GARCH specifications, a subclass is very similar to the usual formulations. For $\alpha_i^+ = \alpha_i^- = \alpha_i$ for all i, the conditional standard deviation is

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^{p} \beta_j \sigma_{t-j}, \tag{7}$$

which provides very close trajectories to those obtained with the standard GARCH. In this case, the volatility is only function of the magnitude of the perturbations.[2]

The central feature of the model is that it allows different reactions of the current volatility to the sign of past innovations. The effect of a shock ε_{t-k} on σ_t is a function of both its magnitude and its sign. For example, take the Threshold ARCH(1) with an assumption of symmetric distribution for (Z_t) and suppose that the second-order moments of (ε_t) exist (the conditions are given below). An easy computation (see appendix) proves that $\text{cov}(\varepsilon_{t-k}, \sigma_t)$ is proportional to $(\alpha_1^+ - \alpha_1^-)$ with a positive constant of proportionality. In particular it is null in the symmetric case (7).

Exponential and Threshold GARCH have some similarities since they both model asymmetries in volatility. It seems important to point out their differences. First, the latter provides an additive modelling and makes volatility a function of (nonnormalized) innovations, which the former does not, as noted in the introduction. For this reason, it is actually closer to the classical formulations. The second difference is in the modelling of asymmetries: exponential

[1] As in GARCH models the constraints can be weakened [Nelson and Cao (1992)].

[2] Model (7) has been applied by Taylor (1986) and by Schwert (1989).

GARCH impose a constant structure at all lags. In the threshold GARCH case, different lags may yield opposite contributions as far as asymmetry is concerned (for example, $\alpha_1^+ - \alpha_1^- > 0$ while $\alpha_2^+ - \alpha_2^- < 0$). This need of complexity in the asymmetry structure will be illustrated in section 4. Finally, exponential GARCH modelling implies an ARMA equation for the $\ln\sigma^2$ process, which makes the probabilistic study very easy. However, it does not provide any linear equation in any function of ε. As a consequence, two-step least squares inference procedures are not possible in the exponential GARCH context. Threshold GARCH preserves this feature of classical GARCH models, as shown in the following section.

2.1. *Linearity of threshold GARCH*

From (3) and (6) we have

$$\varepsilon_t^+ = \sigma_t Z_t^+ \quad \text{and} \quad \varepsilon_t^- = \sigma_t Z_t^- . \tag{8}$$

Then, using (5), the conditional means of the positive and negative parts are

$$E[\varepsilon_t^+ / \underline{\varepsilon_{t-1}}] = \sigma_t E[Z^+] \quad \text{and} \quad E[\varepsilon_t^- / \underline{\varepsilon_{t-1}}] = \sigma_t E[Z^-] .$$

Let u_t and v_t denote the respective innovations of ε_t^+ and ε_t^-; we have

$$\varepsilon_t^+ - u_t = \sigma_t E[Z^+] \quad \text{and} \quad \varepsilon_t^- - v_t = \sigma_t E[Z^-] .$$

Multiplying (4) by $E[Z^+]$ first and then by $E[Z^-]$ yields

$$\varepsilon_t^+ = \alpha_0 E[Z^+] + \sum_{i=1}^{q} [\alpha_i^- \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^-] E[Z^+] + \sum_{j=1}^{p} \beta_j (\varepsilon_{t-j}^+ - u_{t-j}) + u_t ,$$

$$\varepsilon_t^- = \alpha_0 E[Z^-] + \sum_{i=1}^{q} [\alpha_i^- \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^-] E[Z^-] + \sum_{j=1}^{p} \beta_j (\varepsilon_{t-j}^- - v_{t-j}) + v_t . \tag{9}$$

Therefore, the vector $(\varepsilon_t^+, \varepsilon_t^-)$ is solution of an ARMA$(\max(p,q), p)$ equation. This is analogous to the GARCH case in which the same property is true for the (ε_t^2) process.

To evaluate whether shocks to variance persist or not, it is necessary to analyse the stationarity properties of the error process. As in the GARCH context, a general study is difficult. We concentrate on two important cases: the threshold ARCH(q) and the threshold GARCH$(1, 1)$.

2.2. Threshold ARCH

In that case, volatility is a function of the past innovations only,

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^+ \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^- . \tag{10}$$

In order to analyse the second-order stationarity properties of the (ε_t) process, we have to compute several moments: $E\varepsilon_t$, $E\varepsilon_t^2$, $E(\varepsilon_t\varepsilon_{t-h})$. To achieve this we introduce the following state-vector:

$$\omega_t = [\varepsilon_t^{+2}, \varepsilon_t^{-2}, \varepsilon_t^+\varepsilon_{t-1}^+, \varepsilon_t^-\varepsilon_{t-1}^+, \varepsilon_t^+\varepsilon_{t-1}^-, \varepsilon_t^-\varepsilon_{t-1}^-, \ldots, \varepsilon_t^+\varepsilon_{t-q+1}^+,$$

$$\varepsilon_t^-\varepsilon_{t-q+1}^+, \varepsilon_t^+\varepsilon_{t-q+1}^-, \varepsilon_t^-\varepsilon_{t-q+1}^-, \varepsilon_t^+, \varepsilon_t^-]' .$$

We prove that (ω_t) has an AR(q) representation:

Proposition 1. There exist q square matrices A_i and a vector b such that

$$\omega_t = b + \sum_{i=1}^{q} A_i\omega_{t-i} + \eta_t \quad \textit{with} \quad E[\eta_t/\underline{\omega_{t-1}}] = 0 .$$

Proof. See appendix.

Let $A(L)$ denote the lags polynomial $A(L) = I - \sum_{i=1}^{q} A_i L^i$ and $u' = (1, 1, \ldots, 0) \in \mathrm{R}^q$.

The next property gives a sufficient condition for wide-sense stationarity:

Proposition 2. Suppose that $\det A(z) = 0 \Rightarrow |z| > 1$. *Then the Threshold ARCH(q) process as defined in (3), (5), and (10) is wide-sense stationary, with*

$$E\varepsilon_t = 0, \quad V(\varepsilon_t) = u'[A(1)]^{-1}b, \quad \operatorname{cov}(\varepsilon_t, \varepsilon_s) = 0 \quad \textit{for} \quad t \neq s.$$

Proof. See appendix.

A similar, but tedious, analysis may be conducted in the general case by including past values of σ_t (and their cross-products with the other variables) in ω_t.

2.3. Threshold GARCH(1, 1)

The simplest (but useful for many applications) Threshold GARCH model sets

$$\sigma_t = \alpha_0 + \alpha_1^+\varepsilon_{t-1}^+ - \alpha_1^-\varepsilon_{t-1}^- + \beta\sigma_{t-1}, \quad \alpha_0 > 0, \quad \alpha_1^+ \geq 0, \quad \alpha_1^- \geq 0. \tag{11}$$

We rewrite (11) in order to have an AR(1) equation for σ_t, with a random coefficient depending on Z,

$$\sigma_t = \alpha_0 + B(Z_{t-1})\sigma_{t-1}\,, \tag{12}$$

with $B(Z_{t-1}) = \alpha_1^+ Z_{t-1}^+ - \alpha_1^- Z_{t-1}^- + \beta$. This relation is used to investigate the problems of strict and second-order stationarity.

To obtain the strict stationarity condition, we suppose that $E[\log B(Z_t)]$ exists. Following the same approach as Nelson (1990), we prove that the existence of a strict-stationary solution for the Threshold GARCH(1) model depends on the sign of this quantity.

Proposition 3. *Let $E[\log B(Z_t)]$ be nonzero. Then there exists a strict-stationary Threshold GARCH(1, 1) solution if and only if $E[\log B(Z_t)] < 0$. If there is a solution it is unique.*

Proof. See appendix.

When there is no drift in the volatility, it converges either to zero or to infinity.

Proposition 4. *Suppose $\alpha_0 = 0$. Then:*

(i) *if $E[\log B(Z_t)] < 0$, σ_t goes to zero almost surely.*

(ii) *if $E[\log B(Z_t)] > 0$, σ_t goes to infinity almost surely.*

Proof. See appendix.

In this simple case it is also possible to get a necessary and sufficient weak-stationarity condition.

Proposition 5. *A necessary and sufficient condition for wide-sense stationarity is: $E[B(Z_t)]^2 < 1$. The variance is then given by*

$$V(\varepsilon_t) = \alpha_0^2[1 + E[B(Z_t)][1 - E[B(Z_t)]]^{-1}[1 - E[B(Z_t)]^2]^{-1}\,.$$

Furthermore, the weak-stationary solution is unique when it exists.

Proof. See appendix.

The following property is analogous to the one obtained in the GARCH case.

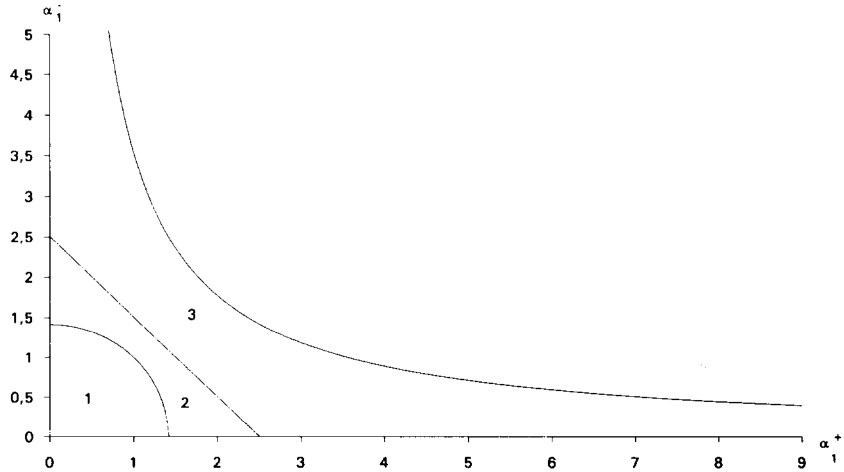

Fig. 1. Stationarity conditions for the Threshold ARCH(1) (plotted under the normality assumption). *Region 1:* Weak stationarity, existence of $E\varepsilon^+$, $E\varepsilon^-$, $E\sigma$, strict stationarity. *Region 2:* Existence of $E\varepsilon^+$, $E\varepsilon^-$, $E\sigma$, strict stationarity. *Region 3:* Strict stationarity.

Proposition 6. When there is a weak-stationary solution it is stationary in the strict sense.

Proof. See appendix.

When $\beta_1 = 0$ and the distribution of (Z_t) is *symmetric*, one can easily obtain the following forms for the strict- and weak-stationarity conditions, respectively:

$$\alpha_1^+ \alpha_1^- < \exp[-2E\log|Z_t|] \quad \text{and} \quad \alpha_1^{+2} + \alpha_1^{-2} < 2.$$[3]

We finally obtain a quadrant, a segment of line, and a branch of hyperbole as respective boundaries for the regions of weak stationarity, existence of expectations for $\varepsilon_t^+, \varepsilon_t^-, \sigma_t$, and strict stationarity in the Threshold ARCH(1) case.

Higher-order moments may be analyzed in a similar fashion, supposing that the Z process has moments to the appropriate order. The next proposition gives conditions for their existence and a recursive way to compute them.

[3]When the distribution of (Z_t) is standard Gaussian [i.e., (ε_t) conditionally normal], we have $E[\log|Z_t|] = (1/2)[\log 2 + \psi(1/2)]$, where ψ denotes the Euler function [see Nelson (1991), Abramowitz and Stegun (1964)]. Therefore the strict-stationarity condition writes: $\alpha_1^+ \alpha_1^- < (1/2)\exp[-\psi(1/2)] \approx 3.5619$.

Proposition 7. For the Threshold ARCH(1) process, a necessary and sufficient condition for existence of the kth moment is: $E[B(Z_t)]^k < 1$. *The kth moment can be expressed by the recursive formula*

$$E(\varepsilon_t^k) = E(\sigma_t^k)E(Z_t^k) \quad \textit{with} \quad E(\sigma_t^k) = \sum_{i=0}^{k} \binom{k}{i} \alpha_0^i E[B(Z_t)]^{k-i} E(\sigma_t)^{k-i}.$$

Proof. See appendix.

Note that in the Threshold ARCH(1) case $E[B(Z_t)]^k = (\alpha_1^+)^k E[Z_t^+]^k + (-\alpha_1^-)^k E[Z_t^-]^k$, and that if the distribution of (Z_t) is *symmetric* $E(\varepsilon_t^{2h-1}) = 0$ for all $h > 0$.

It is well known that the GARCH(p,q) process (with Gaussian conditional distribution) generates data with fatter tails than the normal density [see Engle (1982), Milhoj (1984)]. The same is true with conditionally Gaussian Threshold GARCH.

Let us, for example, consider the constraints for the existence of moments in the Threshold ARCH(1) case: $[(\alpha_1^+)^k + (\alpha_1^-)^k]^{1/k} < [\|Z^+\|_k]^{-1}$. Using the Stirling formula, it is easy to show that the equivalent of $\|Z^+\|_k$ for large k is in $k^{1/2}$. Hence we have $\max(\alpha_1^+, \alpha_1^-) \le [(\alpha_1^+)^k + (\alpha_1^-)^k]^{1/k} \le 0$ if all the moments exist, which proves that only the degenerate process has finite moments of arbitrary order. Another way to highlight the property of fat tails is by considering the Kurtosis coefficient, ratio of the L_4 and the squared L_2 norms. If we note $\mathscr{K}_Z$ and $\mathscr{K}_\varepsilon$ the Kurtosis coefficients of Z and ε, we have

$$\mathscr{K}_\varepsilon = \mathscr{K}_Z \cdot E[\sigma_t^4]/(E[\sigma_t^2])^2 > \mathscr{K}_Z,$$

which proves leptokurticity. Simple computations in the symmetric case $(\alpha_1^+ = \alpha_1^-)$ show that the coefficient of Kurtosis goes to infinity when the parameter goes to the constraint relative to the fourth moment.

3. Estimation of autoregressive models with Threshold GARCH errors

In this section we consider the simple AR(1) process (Y_t) solution of

$$Y_t = \varphi Y_{t-1} + \varepsilon_t, \qquad |\varphi| < 1, \tag{13}$$

where (ε_t) is given by (3), (4), and (5), with the additional assumption that the (Z_t) process has a *normal distribution*. We also suppose that the wide-sense stationarity conditions are satisfied. Of course, more general regression models might be considered as well. However, our model seems to be sufficient for many financial applications where a somewhat simple but standard assumption is a random walk for the logarithms of prices (corresponding to $\varphi = 0$, if Y_t is computed as the logarithm of the return). The first problem that we consider is the variance prediction. For simplicity, we restrict ourself to the Threshold ARCH case.

3.1. Prediction properties of the Threshold ARCH(q)

The model handles prediction problems just as the usual AR(1) model, with

$$E[Y_{t+s}/\underline{Y_t}] = \varphi^s Y_t, \qquad s \geq 0.$$

Obviously, the modification is in the precision of the prediction. We have:

$$V[Y_{t+s}/\underline{Y_t}] = E\{[Y_{t+s} - E[Y_{t+s}/\underline{Y_t}]]^2/\underline{Y_t}\} = \sum_{i=1}^{s} \varphi^{2(s-i)} E[\varepsilon_{t+i}^2/\underline{\varepsilon_t}].$$

Using the AR(q) representation of Proposition 1, it is easy to see that the prediction of the state-vector has the following form:

$$E[\omega_t/\underline{\omega_{t-s}}] = b^* + \sum_{i=1}^{q} A_i^* \omega_{t-s-i+1},$$

where b^* and A_i^* depend on vector b and matrices A_1.

Since ε_t^{+2} and ε_t^{-2} (first components of ω_t) have the same predictions, the conditional precision deduces from a function of the form

$$V[Y_{t+s}/\underline{Y_t}] = c_0 + \sum_{i=1}^{q} c_i' \omega_{t-i},$$

the real c_0 and the vectors c_i depending on φ, b, and A_i, $i = 1, q$.

Hence the precision depends on the past through the q lagged values of ω_t. As expected this dependence is both on the size and on the sign of past innovations. Simple computations in the case $q = 1$ prove that the asymptotic precision (when s goes to infinity) is independent of the past.

3.2. Maximum likelihood estimation

Let $\omega' = [\alpha_0, \alpha_1^+, \alpha_1^-, \ldots, \beta_{p-1}, \beta_p]$ and $\theta = [\varphi, \omega'] \in \Theta$, compact subspace of $\mathbb{R}^{2q+p+1}$. The loglikelihood function for a sample of T observations, conditional on the first q, is apart from some constant:

$$\log L_T(Y; \theta) = -\sum_{t=q+1}^{T} \log \sigma_t - \tfrac{1}{2} \sum_{t=q+1}^{T} (\varepsilon_t^2/\sigma_t^2). \tag{14}$$

Compared to the GARCH case, it is important to note that this function is continuous in θ, differentiable with respect to ω but not always with respect to φ because of the presence of thresholds. The maximum likelihood estimator $\hat{\theta}_T$ of parameter θ is then a particular case of M-estimator based on a continuous, right differentiable function. We know that under certain regularity conditions [Huber (1967), Gouriéroux and Monfort (1989, ch. 8)] this estimator is consistent, asymptotically normal:

$$\sqrt{T}(\hat{\theta}_T - \theta_T) \xrightarrow{d} \mathcal{N}\,[0, J^{-1}]\,, \tag{15}$$

with

$$J = E_0\left[\frac{\partial^+ \log \ell(Y;\theta_0)}{\partial\theta}\frac{\partial^+ \log \ell(Y;\theta_0)}{\partial\theta'}\right] = \left[\frac{\partial}{\partial\theta}E_0\left[\frac{\partial^+ \log \ell(Y;\theta_0)}{\partial\theta}\right]\right]_{\theta=\theta_0},$$

where $\partial^+/\partial\theta$ denotes the right derivation, component by component, and $\ell(Y;\theta)$ the density function of Y_t conditional on $\underline{Y_{t-1}}$.

Checking the regularity conditions ensuring consistency and asymptotic normality has proved very difficult in GARCH models [see Weiss (1984)]. This problem is also far from being solved in our case. Therefore, asymptotic normality is considered as a conjecture for the remainder of this paper.

Note that in the loglikelihood, the problem of nondifferentiability arises only for σ_t and with respect to φ. Differentiating with respect to the mean and variance parameters yields

$$\frac{\partial^+ \log L_T(Y;\theta_T)}{\partial\varphi} = \sum_{t=q+1}^{T}\frac{1}{\sigma_t^3}\left[(\varepsilon_t^2 - \sigma_t^2)\frac{\partial^+\sigma_t}{\partial\varphi} - \sigma_t\varepsilon_t\frac{\partial\varepsilon_t}{\partial\varphi}\right].$$

$$\frac{\partial \log L_T(Y;\theta_T)}{\partial\omega} = \sum_{t=q+1}^{T}\frac{1}{\sigma_t^3}\frac{\partial\sigma_t}{\partial\omega}[\varepsilon_t^2 - \sigma_t^2].$$

At the optimum we then have

$$\sum_{t=q+1}^{T}\frac{1}{\hat{\sigma}_t^3}\frac{\partial\hat{\sigma}_t}{\partial\alpha}[\hat{\varepsilon}_t^2 - \hat{\sigma}_t^2] = 0\,,$$

which may be seen as an orthogonality condition between the residual corresponding to the squared error (ε_t^2 diminished of its conditional mean σ_t^2) and some implicit explanatory variables.

To obtain maximum likelihood estimates, a possible method is to replace the function $x^+ = \max(x, 0)$ by a differentiable approximation and apply the usual method based on the first-order conditions. But better is certainly to apply the classical algorithms, leaving the software derive numerically.

It follows from (15) that the asymptotic precision deduces from matrix J, which may be decomposed into blocks:

$$J = \begin{bmatrix} J_{\varphi\varphi} & J_{\varphi\theta} \\ J_{\theta\varphi} & J_{\theta\theta} \end{bmatrix},$$

with

$$J_{\theta\theta} = 3\,E_0\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t}{\partial\theta}\frac{\partial\sigma_t}{\partial\theta'}\right],$$

$$J_{\varphi\varphi} = E_0\left[\frac{1}{\sigma_t^2}\left(\frac{\partial\varepsilon_t}{\partial\varphi}\right)^2 + 3\frac{1}{\sigma_t^2}\left(\frac{\partial^+\sigma_t}{\partial\varphi}\right)^2\right],$$

$$J_{\theta\varphi} = 3\,E_0\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t}{\partial\theta}\frac{\partial^+\sigma_t}{\partial\varphi}\right].$$

Here, the difference with the GARCH(q) regression model is the fact that the second diagonal element $J_{\theta\varphi}$ is generally nonzero. Hence, the mean and variance parameters are linked and inferences on both types of parameters cannot be carried out separately. Again in the symmetric case, the estimators turn out to be asymptotically independent.

Proposition 8. For the Threshold GARCH (p, q) process with $\alpha_i^+ = \alpha_i^-$, $i = 1, q$, *the estimators* $\hat{\varphi}_T$ *and* $\hat{\theta}_T$ *are asymptotically independent.*

Proof. See appendix.

3.3. Least squares estimation of the Threshold ARCH(q)

Because of linearity (described in section 2), a two-step method may also be considered in the Threshold ARCH case [eq. (10)]. Applying O.L.S. to eq. (13) yields the usual estimator $\tilde{\varphi}_T$ and the residuals $\tilde{\varepsilon}_t = Y_t - \tilde{\varphi}_T Y_{t-1}$.

From eq. (9) we get

$$\varepsilon_t^+ = \left[\alpha_0 + \sum_{i=1}^{q}[\alpha_i^+\varepsilon_{t-i}^+ - \alpha_i^-\varepsilon_{t-i}^-]\right]EZ_t^+ + \sigma_t[Z_t^+ - EZ_t^+],$$

$$\varepsilon_t^- = \left[\alpha_0 + \sum_{i=1}^{q}[\alpha_i^+\varepsilon_{t-i}^+ - \alpha_i^-\varepsilon_{t-i}^-]\right]EZ_t^- + \sigma_t[Z_t^- - EZ_t^-]. \qquad (16)$$

Then we obtain consistent estimators, say $\tilde{\alpha}_T$, of the variance parameters α, by regressing $(\tilde{\varepsilon}_t^+, \tilde{\varepsilon}_t^-)$ on its q past values and a vector of ones.

In a second step, the mean parameter may be improved taking into account the form of the variance of the error term ε_t. This variance may be approached by

$$\tilde{\sigma}_t^2 = \left[\tilde{\alpha}_{OT} + \sum_{i=1}^{q} [\tilde{\alpha}_{iT}^+ \tilde{\varepsilon}_{t-i}^+ - \tilde{\alpha}_{iT}^- \tilde{\varepsilon}_{t-i}^-] \right]^2 .$$

Hence a quasi-generalized least square estimator of φ,

$$\tilde{\tilde{\varphi}}_T = \left[\sum_{t=2}^{T} \frac{Y_t Y_{t-1}}{\tilde{\sigma}_t^2} \right] \left[\sum_{t=2}^{T} \frac{Y_{t-1}^2}{\tilde{\sigma}_t^2} \right]^{-1},$$

with asymptotic variance

$$V_{\mathrm{as}}[\sqrt{T}(\tilde{\tilde{\varphi}}_T - \varphi)] = \left[E_0 \left[\frac{1}{\sigma_t^2} \left(\frac{\partial \varepsilon_t}{\partial \varphi} \right)^2 \right] \right]^{-1} .$$

A comparison with the term $J_{\varphi\varphi}$ in the information matrix shows that, even in the favourable case where the two inferences on φ and α may be conducted separately, the Q.G.L.S. estimator is not asymptotically efficient.

Taking into account the covariance structure of the error terms in eq. (16), the estimators of the variance parameters may also be improved. For two different indexes these errors terms are uncorrelated. Finally, a Q.G.L.S. estimator of α may be obtained by regressing $(\tilde{\varepsilon}_t^+, \tilde{\varepsilon}_t^-)$ on its q past values and a vector of ones, with covariance structure:

$$\Omega = \operatorname{diag}(\tilde{h}_t^2) \otimes \begin{bmatrix} EZ_t^{+2} - (EZ_t^+)^2 & (EZ_t^+)(EZ_t^-) \\ (EZ_t^+)(EZ_t^-) & EZ_t^{-2} - (EZ_t^-)^2 \end{bmatrix} .$$

4. Example of the French CAC stock index

In this section, the Threshold GARCH approach is compared with the classical and exponential GARCH, using daily stock-index data.

The set of data consists of a series of the French CAC stock index, built with 240 assets. It contains 3606 daily returns covering the period from January '76 to July '90. The whole period was partitioned in three subperiods of about five years each. Let $Y_t = \ln(X_t/X_{t-1})$, where X_t is the CAC value at date t.

From a preliminary empirical analysis, an AR(1) model was selected for the conditional mean:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \varepsilon_t . \tag{17}$$

We first estimated an ARCH process with a short memory. A five-lags structure was selected for the conditional variances:

$$\text{ARCH(5):} \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{5} \alpha_i \varepsilon_{t-i}^2 .$$

The Berndt, Hall, Hall, and Hausman (1974) algorithm was used to obtain the estimates. The gradient was computed numerically. Parameter estimates are insensitive to various initial conditions for our sample, making it likely that global maxima are achieved. The estimation results are reported in table 1.

First note that the sets of coefficient estimates are quite different from one period to another. All of them are statistically significant (except Φ_0 in period 1 and α_4 in period 2). The sum of the ARCH parameters ($\alpha_1 + \cdots + \alpha_5$) is substantially smaller than unity. In every subperiod, conditional homoskedasticity is rejected.

To catch nonsymmetric behaviours in the conditional variances we then considered a Threshold ARCH specification with the same length in the lag structure:

$$\text{TARCH(5):} \qquad \sigma_t = \alpha_0 + \sum_{i=1}^{5} \alpha_i^+ \varepsilon_{t-i} - \alpha_i^- \varepsilon_{t-i}^- . \tag{18}$$

We have reported the estimates in table 2.

Once again, most of the estimates are statistically significant; moreover, at least one of the positive and negative parts of ε_{t-i} has a significant coefficient for each lag and each subperiod. To compute the standard deviations estimates of process Y, we used the AR(5) representation of Proposition 1, the matrices being replaced by their estimates. In each period the conditions of Proposition 2 are satisfied, proving that the fitted processes are at least second-order stationary.

The asymmetric reaction of volatility to the sign of past shocks is highly significant at each period. To check it we have done Wald tests on the differences $\alpha_i^+ - \alpha_i^-$, using the asymptotic covariance matrix of the parameter estimates. The conclusion was that at least three of these differences were found to be significantly different from zero for each sample. Do these results support the hypothesis that volatility is higher after a decrease than after an increase? The answer is far from being evident as each subperiod provides at least one significant $\alpha_i^+ - \alpha_i^-$ of each sign. To go further in the analysis, let us interpret the asymmetry as a correlation between past shocks and the current volatility, and compute the estimates of $\text{corr}(\sigma_t, \varepsilon_{t-k})$, $k = 1, 5$ (recall that these correlations are null in the ARCH case as well as in the symmetric one).

Table 1
ARCH(5) parameter estimates (standard errors in parentheses).

Parameter	1976–80	1981–85	1986–90	1976–90
Φ_0	0.01[a]	0.08	0.09	0.06
	(0.02)	(0.02)	(0.03)	(0.01)
Φ_1	0.33	0.33	0.11	0.27
	(0.03)	(0.03)	(0.00)	(0.02)
α_0	0.29	0.26	0.48	0.33
	(0.02)	(0.02)	(0.04)	(0.01)
α_1	0.19	0.16	0.16	0.18
	(0.03)	(0.04)	(0.03)	(0.02)
α_2	0.07	0.20	0.25	0.17
	(0.03)	(0.03)	(0.03)	(0.02)
α_3	0.17	0.11	0.05	0.11
	(0.03)	(0.03)	(0.03)	(0.02)
α_4	0.16	0.03[a]	0.18	0.12
	(0.03)	(0.03)	(0.03)	(0.02)
α_5	0.14	0.10	0.06	0.11
	(0.03)	(0.03)	(0.03)	(0.02)
Unconditional std. dev. of Y	1.09	0.86	1.24	1.07

[a]Not significantly different from zero at 5% level.

Table 2
TARCH(5) parameter estimates (standard errors in parentheses).

Parameter	1976–80		1981–85		1986–90		1976–90	
m	− 0.02[a]		0.06		0.05[a]		0.02[a]	
	(0.02)		(0.02)		(0.04)		(0.01)	
φ	0.35		0.32		0.12		0.28	
	(0.03)		(0.03)		(0.03)		(0.02)	
α_0	0.38		0.35		0.54		0.43	
	(0.03)		(0.02)		(0.04)		(0.01)	
α_1^+, α_1^-	0.08	0.25	0.31	0.19	0.29	0.21	0.24	0.24
	(0.02)	(0.03)	(0.00)	(0.03)	(0.00)	(0.04)	(0.02)	(0.00)
α_2^+, α_2^-	0.23	0.24	0.03	0.07	0.11	0.24	0.14	0.15
	(0.00)	(0.04)	(0.00)	(0.00)	(0.04)	(0.00)	(0.02)	(0.02)
α_3^+, α_3^-	0.08	0.22	0.20	0.19	0.05[a]	0.13	0.09	0.18
	(0.00)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.00)
α_4^+, α_4^-	0.17	0.00[a]	0.11	0.08	0.00[a]	0.17	0.07	0.12
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.00)	(0.02)
α_5^+, α_5^-	0.09	0.19	0.15	0.06	0.06	0.13	0.11	0.10
	(0.03)	(0.00)	(0.04)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)
Unconditional std. dev. of Y	1.31		0.99		1.41		1.23	

[a]Not significantly different from zero at 5% level.

Once again, the estimate components of $E\omega_t$ (see Proposition 1) can be used, as

$$\operatorname{cov}(\sigma_t, \varepsilon_{t-k}) = E\sigma_t \varepsilon_{t-k} = \sum_{i=1}^{5} \alpha_i^+ (E\varepsilon_{t-k}^+ \varepsilon_{t-i}^+ + E\varepsilon_{t-k}^- \varepsilon_{t-i}^+)$$

$$- \alpha_i^- (E\varepsilon_{t-k}^+ \varepsilon_{t-i}^- + E\varepsilon_{t-k}^- \varepsilon_{t-i}^-) .$$

Table 3 gives these correlations. The standard deviations are computed using the variance–covariance matrix of the parameters estimates.

In table 3, the majority of the estimated correlations appear significant. Note that the signs of the correlations do not systematically duplicate those of the differences $\alpha_i^+ - \alpha_i^-$, ε_{t-i} being correlated with ε_{t-i}^+ and ε_{t-i}^-, $0 < i < k$ (for example in period 3, $\alpha_2^+ - \alpha_2^- < 0$, while the corresponding estimated correlation is positive). More important, these results do not confirm evidence that the current volatility is negatively correlated with past innovations: only six of the thirteen significant correlations are negative according to table 3.

It is clear from the previous results that a longer memory in the conditional variance is called for. Therefore, we extend the comparison to models of the GARCH(1, 1) type. Four models are estimated over the different subperiods: the classical GARCH, the threshold GARCH, the exponential GARCH, and the GARCH with absolute values instead of squares. The latter is a particular case of the threshold model, in which the coefficients corresponding to the positive

Table 3

Estimated correlations between σ_t and ε_{t-k}, $k = 1, 5$ (standard errors in parentheses).

k	1976–80	1981–85	1986–90	1976–90
1	− 0.15	0.12	0.08	0.00[a]
	(0.03)	(0.03)	(0.04)	(0.01)
2	− 0.03[a]	− 0.02	0.14	− 0.00[a]
	(0.03)	(0.00)	(0.00)	(0.02)
3	− 0.17	0.02[a]	− 0.05[a]	− 0.10
	(0.03)	(0.06)	(0.06)	(0.02)
4	0.09	0.06[a]	− 0.15	− 0.07
	(0.04)	(0.05)	(0.05)	(0.02)
5	− 0.02[a]	0.35	0.15	0.19
	(0.03)	(0.05)	(0.05)	(0.03)

[a]Not significantly different from zero at 5% level.

Table 4

Parameter estimates for the CAC 240, 1976–1990 (standard errors in parentheses).

Parameter	Model			
	GARCH(1, 1)	STGARCH(1, 1)	TGARCH(1, 1)	EGARCH(1, 1)
Φ_0	0.056 (0.013)	0.025[a] (0.013)	0.038 (0.013)	0.039 (0.013)
Φ_1	0.262 (0.017)	0.265 (0.019)	0.262 (0.017)	0.272 (0.017)
α_0	0.042 (0.007)	0.043 (0.012)	0.049 (0.017)	− 0.225 (0.012)
α_1 or α_1^+			0.111 (0.020)	0.238 (0.018)
α_1 or α_1^-	0.156 (0.009)	0.155 (0.015)	0.192 (0.019)	0.343 (0.016)
β_1	0.807 (0.011)	0.836 (0.018)	0.833 (0.020)	0.958 (0.004)

[a]Not significantly different from zero at 5% level.

and negative parts are equal: we call it symmetric threshold GARCH. The corresponding volatilities are

$$\text{GARCH:} \quad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 ,$$

$$\text{TGARCH:} \quad \sigma_t = \alpha_0 + \alpha_1^+ \varepsilon_{t-1}^+ - \alpha_1^- \varepsilon_{t-1}^- + \beta\sigma_{t-1} ,$$

$$\text{EGARCH:} \quad \ln\sigma_t^2 = \alpha_0 + \alpha_1^+ Z_{t-1}^+ - \alpha_1^- Z_{t-1}^- + \beta \ln\sigma_{t-1}^2 ,$$

$$\text{STGARCH:} \quad \sigma_t = \alpha_0 + \alpha_1 |\varepsilon_{t-1}| + \beta\sigma_{t-1} .$$

Tables 4–7 give the parameter estimates of the models.

First note that the mean parameters estimates are nearly identical for each model and each period. Once again we find strong evidence of asymmetry in the data, except for the 1981–1986 period. For all other periods, both exponential and threshold GARCH model give evidence of asymmetry in the same direction: past negative returns have a higher impact on the current volatility than positive ones.

Although the models are nonnested, it is possible to use the normal quasi-likelihood to compare the different specifications: Nelson (1990b) has given a theoretical support to this procedure.[4] According to table 8, this selection

[4]Although his definition of 'noisy' variance estimates cannot be directly applied to the normalized residuals from estimated ARCH models.

Table 5
Parameter estimates for the CAC 240, 1976–1980 (standard errors in parentheses).

Parameter		Model: GARCH(1, 1)	STGARCH(1, 1)	TGARCH(1, 1)	EGARCH(1, 1)
Φ_0		0.010[a]	0.025[a]	– 0.007[a]	– 0.004[a]
		(0.022)	(0.021)	(0.021)	(0.022)
Φ_1		0.337	0.331	0.328	0.331
		(0.031)	(0.026)	(0.029)	(0.029)
α_0		0.042	0.045	0.048	– 0.240
		(0.007)	(0.014)	(0.017)	(0.020)
α_1 or	α_1^+	0.176	0.166	0.113	0.219
		(0.016)	(0.021)	(0.035)	(0.033)
	α_1^-			0.205	0.376
				(0.032)	(0.028)
β_1		0.791	0.824	0.829	0.943
		(0.016)	(0.029)	(0.029)	(0.009)

[a]Not significantly different from zero at 5% level.

Table 6
Parameter estimates for the CAC 240, 1981–1985 (standard errors in parentheses).

Parameter		Model: GARCH(1, 1)	STGARCH(1, 1)	TGARCH(1, 1)	EGARCH(1, 1)
Φ_0		0.071	0.065	0.059	0.061
		(0.019)	(0.017)	(0.019)	(0.019)
Φ_1		0.329	0.326	0.324	0.353
		(0.030)	(0.028)	(0.028)	(0.031)
α_0		0.045	0.051	0.056	– 0.260
		(0.007)	(0.017)	(0.019)	(0.025)
α_1 or	α_1^+	0.143	0.167	0.133	0.326
		(0.019)	(0.026)	(0.042)	(0.041)
	α_1^-			0.179	0.336
				(0.037)	(0.029)
β_1		0.784	0.794	0.806	0.965
		(0.025)	(0.038)	(0.045)	(0.006)

criterion leads to prefer asymmetric formulations rather than classical ones at all periods, the related loglikelihood being always higher for the former than for the latter. The threshold GARCH seems to be preferable than the exponential GARCH for the first and third subperiods, while the converse is true for the

Table 7

Parameter estimates for the CAC 240, 1986–1990 (standard errors in parentheses).

Parameter	Model			
	GARCH(1, 1)	STGARCH(1, 1)	TGARCH(1, 1)	EGARCH(1, 1)
Φ_0	0.096 (0.029)	0.063 (0.024)	0.063 (0.029)	0.068 (0.029)
Φ_1	0.103 (0.031)	0.121 (0.027)	0.108 (0.029)	0.101 (0.033)
α_0	0.071 (0.017)	0.058 (0.025)	0.070 (0.029)	− 0.174 (0.020)
α_1 or α_1^+	0.156 (0.018)	0.134 (0.032)	0.081 (0.036)	0.157 (0.033)
α_1 or α_1^-			0.193 (0.038)	0.314 (0.028)
β_1	0.797 (0.027)	0.847 (0.036)	0.831 (0.044)	0.946 (0.012)

Table 8

Likelihood ratios statistics.

Model	LR[a]			
	1976–80	1981–85	1986–90	1976–90
ARCH (5)	119.2	158.2	180.0	469.0
TARCH (5)	127.6	169.7	196.2	486.5
GARCH (1, 1)	104.4	145.5	164.8	486.0
STGARCH (1, 1)	107.2	144.6	165.2	490.4
TGARCH (1, 1)	112.3	151.5	178.9	501.2
EGARCH (1, 1)	105.3	210.3	165.7	513.9

[a] *LR* is the excess loglikelihood compared with the homoskedastic model.

other periods. Apart from asymmetry, table 8 also shows that it seems better to model volatility through standard deviations rather than through variances (although we do not know if the differences are significant).

Another way to assess the quality of the results is to consider the sample excess Kurtosis for the standardized residuals $\hat{\varepsilon}_t/\hat{\sigma}_t$. Once again we refer to Nelson (1990b) (with the same comments as before), who showed that 'the greater the error in the volatility estimate, the thicker-tailed are the standardized residuals'. As documented by many previous studies, the models do not fully account for the observed leptokurtosis in the series. However, the conclusions

Table 9

Estimated kurtosis coefficient for the normalized residuals.

$\mathscr{K}$	1976–80	1981–85	1986–90	1976–90
ARCH(5)	3.32	2.86	1.61	2.81
TARCH(5)	2.64	1.95	1.19	2.29
GARCH(1, 1)	3.16	3.26	1.97	2.98
STGARCH(1, 1)	3.07	3.19	1.95	2.89
TGARCH(1, 1)	2.83	2.78	1.48	2.48
EGARCH(1, 1)	2.83	2.75	1.55	2.76

confirm the previous ones. The improvement over traditional models is mainly caused by the use of asymmetric formulations, but it also seems to be due to the modelling of standard deviations. From this example the performances of the two models incorporating asymmetries are very similar over the subperiods. However, it seems that the threshold GARCH takes a better account of large innovations than the exponential GARCH over the whole period. Finally, note that according to this criterion as well as the preceding one, the threshold ARCH with five lags in the volatility does better than all the others. Unfortunately, we have no confidence intervals, hence is it hard to know how significant these results are.

5. Conclusion

The threshold model introduced in this paper should provide an effective tool, simple to use and estimate, for reflecting the asymmetric relation between volatility and past returns. Its simplicity is the one of classical GARCH: the modelling provides additivity, linearity, and dependence of volatility in the innovations magnitude.

Its generality with respect to the treatment of asymmetry is illustrated in the empirical example. Our results confirm recent evidence that volatility of stock returns reacts differently to increases and decreases. However, the evidence of a negative correlation between past innovations and the current volatility is far from being fully supported by our data. Models with one lag in each variable do confirm this hypothesis (threshold or exponential GARCH agree on this point), except for one subperiod. However, threshold ARCH with several lags leads to a more subtle analysis. Mainly, it provides the possibility of a nonconstant structure of asymmetry: the relative impacts of negative and positive shocks on the current volatility are not the same at all lags. Our data typically illustrate this possibility, which leads to moderate the conclusion concerning negative correlations.

Appendix

A.1. Computation of $\operatorname{cov}(\varepsilon_{t-k},\sigma_t)$ *for the threshold ARCH(1)*

$$\begin{aligned}
\operatorname{cov}(\varepsilon_{t-k},\sigma_t) &= E(\varepsilon_{t-k}\sigma_t) = E(\varepsilon_{t-k}(\alpha_1^+\varepsilon_{t-1}^+ - \alpha_1^-\varepsilon_{t-1}^-)) \\
&= E(\varepsilon_{t-k}\sigma_{t-1}(\alpha_1^+ Z_{t-1}^+ - \alpha_1^- Z_{t-1}^-)) \\
&\quad \text{as} \quad \varepsilon_t^+ = \sigma_t Z_t^+ \quad \text{and} \quad \varepsilon_t^- = \sigma_t Z_t^-, \\
&= E(\varepsilon_{t-k}\sigma_{t-1})(\alpha_1^+ + \alpha_1^-)EZ^+ \ \text{from symmetry and} \\
&\quad \text{independence,} \\
&= E(\varepsilon_{t-k}\sigma_{t-k+1})(\alpha_1^+ + \alpha_1^-)^{k-1}EZ^+ \ \text{by iteration,} \\
&= E(\alpha_1^+\varepsilon_{t-k}^{+2} - \alpha_1^-\varepsilon_{t-k}^{-2})(\alpha_1^+ + \alpha_1^-)^{k-1}EZ^+, \\
&= (\alpha_1^+ - \alpha_1^-)(\alpha_1^+ + \alpha_1^-)^{k-1}EZ^+EZ^{+2}E\sigma_{t-k}^2, \qquad k \geq 1.
\end{aligned}$$

A.2. Proof of Proposition 1

The idea is simply to compute the conditional moments of the vector ω_t components, using the fact that $\varepsilon_t^+ = \sigma_t Z_t^+$ and $\varepsilon_t^- = \sigma_t Z_t^-$. For example,

$$E[\varepsilon_t^{+2}/\underline{\varepsilon_{t-1}}] = \sigma_t^2 E[Z_t^{+2}] = \left[\alpha_0 + \sum_{i=1}^{q}\alpha_i^+\varepsilon_{t-i}^+ - \sum_{i=1}^{q}\alpha_i^-\varepsilon_{t-i}^-\right]^2 E[Z_t^{+2}],$$

which may easily, but somewhat tediously, be expressed as a linear function of the components of ω_t. The other expectations are tracted similarly. Q.E.D.

A.3. Proof of Proposition 2

The (ε_t) process being conditionally centred, it is centred and noncorrelated. The condition on the roots of $\det A(z)$ implies the inversibility of $A(1)$. Using the AR(q) model of Proposition 1 and taking expectations yields

$$[E\omega_t - A(1)^{-1}b] = \sum_{i=1}^{q} A_i[E\omega_{t-i} - A(1)^{-1}b].$$

This is a linear recurrence equation for which the stability condition has been imposed. We hence have

$$\lim_{+\infty} E\omega_t = A(1)^{-1}b\,. \tag{A.1}$$

The variance of ε_t follows immediately by $E\varepsilon_t^2 = E\varepsilon_t^{+2} + E\varepsilon_t^{-2}$. Q.E.D.

A.4. Proof of Proposition 3

(i) Sufficient part. Suppose $E[\log B(Z_t)] = \delta < 0$.

Subsequent substitutions in eq. (12) yield

$$\sigma_t = B(Z_{t-1}) \dots B(Z_{t-p})\sigma_{t-p} + \sum_{i=0}^{p-1} B(Z_{t-1}) \dots B(Z_{t-i})\alpha_0 \tag{A.2}$$

[with the convention that $B(Z_{t-1}) \dots B(Z_{t-i}) = 1$ for $i = 0$].

The log $B(Z_t)$ variables being independent, the law of large numbers gives

$$\text{a.s.} \lim \frac{1}{i} \sum_{k=0}^{i-1} \log B(Z_{t-1-k}) = \delta\,.$$

Consequently,

$$\text{a.s.} \lim [B(Z_{t-1}) \dots B(Z_{t-i})]^{1/i} = \mathrm{e}^{\delta} < 1. \tag{A.3}$$

Hence, the existence of

$$\xi_t = \text{a.s.} \lim \sum_{i=0}^{p} B(Z_{t-1}) \dots B(Z_{t-i})\alpha_0\,, \tag{A.4}$$

by application of the Cauchy criterion.

It is easily seen that the (ξ_t) process is solution of eq. (12). $B(Z_{t-1})$ being independent of ξ_{t-1}, (ξ_t) is a Markov process. It is also homogeneous as (Z_t) is strictly stationary, and it is clear from (A.4) that its law is independent of t. It follows that (ξ_t) is strictly stationary.

It remains only to define for all t: $\varepsilon_t = \xi_t Z_t$. The (ε_t) process is solution of eqs. (3) and (11). It is obviously a homogeneous Markov process and the law of ε_t does not depend on t. Hence the existence of a strictly stationary solution.

(ii) Necessary part. Suppose $E[\log B(Z_t)] = \delta > 0$.

From (A.2), we have $\sigma_t \geq \sum_{i=0}^{p-1} B(Z_{t-1}) \ldots B(Z_{t-i})\alpha_0$ for all p. According to (A.3) the series diverges as $e^\delta > 1$.

(iii) Uniqueness of the solution, when $E[\log B(Z_t)] < 0$.

It comes from the fact that $B(Z_{t-1}) \ldots B(Z_{t-p})$ converges to zero in eq. (A.2). Therefore the (ξ_t) process is unique and then (ε_t). Q.E.D.

A.5. Proof of Proposition 4

It follows directly from the previous discussion, using the fact that (A.2) writes $\sigma_t = B(Z_{t-1}) \ldots B(Z_{t-p})\sigma_{t-p}$ in this case. Q.E.D.

A.6. Proof of Proposition 5

(i) Necessary part.

Let (ε_t) be a weak-stationary solution. We have

$$E(\varepsilon_t^2) = E(\sigma_t^2) = \alpha_0^2 + 2\alpha_0 E[B(Z_{t-1})]\, E[\sigma_{t-1}] + E[B(Z_{t-1})]^2 E[\sigma_{t-1}^2]\,.$$

$E[\sigma_t^2]$ being independent of t and all the variables positive, the condition is established.

(ii) Sufficient part. Suppose that $E[B(Z_t)]^2 = \delta < 1$.

We have $E[\log B(Z_t)] \leq \log E[B(Z_t)] \leq \log[E[B(Z_t)]^2]^{1/2}$, using Jensen's inequality.

Then the (ξ_t) process of (A.4) exists. It is clearly stationary in the wide sense as it is noncorrelated and

$$E\xi_t^2 \leq 2\alpha_0^2 E[B(Z_{t-1})^2 + B(Z_{t-1})^2 B(Z_{t-2})^2 + \cdots]$$

$$\leq 2\alpha_0^2 \sum_{k=0}^{+\infty} \delta^k < +\infty.$$

Now we put $\varepsilon_t = \xi_t Z_t$. (ε_t) is a weak-stationary solution of the model and it is also the strict-stationary solution.

(iii) Uniqueness of the solution.

Let $\varepsilon_t^* = \sigma_t^* Z_t$ be another weak-stationary solution process. Then $\sigma_t = \xi_t + \lim_\infty B(Z_{t-1}) \ldots B(Z_{t-p})\, \sigma_{t-p}^*$, the limit being now taken in L^2. We have $E[B(Z_{t-1})^2 \ldots B(Z_{t-p})^2 \sigma_{t-p}^{*2}] = [E[B(Z_{t-1})^2]]^p E\sigma_{t-p}^{*2}$; hence the limit is zero. Q.E.D.

A.7. Proof of Proposition 6

The comparison of the stationary conditions is in section A.5(ii). Q.E.D.

A.8. Proof of Proposition 7

The proof is quite similar to the previous one. The necessary part comes from the expression of $E\sigma_t^k$ deduced from (12) as well as the recursive formula. The sufficient part is based on the existence of a kth moment for the (ξ_t) process which comes from

$$E\xi_t^k \leq k\alpha_0^k E[B(Z_{t-1})^k + B(Z_{t-1})^k B(Z_{t-2})^k + \cdots] \leq k\alpha_0^k \sum_{j=0}^{+\infty} \delta^j < +\infty$$

with $\delta = E[B(Z_t)^k]$. Q.E.D.

A.9. Proof of Proposition 8

We have

$$J_{\alpha\varphi} = 3E_0\left[\frac{1}{\sigma_t^2}[1, |\varepsilon_{t-1}|, \ldots, |\varepsilon_{t-q}|, \ldots, \sigma_{t-p}]' \times \sum_{i=1}^{q} \alpha_i (I_{\varepsilon_{t-i}>0} - I_{\varepsilon_{t-i}<0})(-Y_{t-i-1})\right]$$

$$= 3E_0\left[E_0\left[\frac{1}{\sigma_t^2}[1, |\varepsilon_{t-1}|, \ldots, |\sigma_{t-p}|]' \times \sum_{i=1}^{q} \alpha_i (I_{\varepsilon_{t-i}>0} - I_{\varepsilon_{t-i}<0})(-Y_{t-i-1})/\underline{\varepsilon_{t-2}}\right]\right]$$

$$= 3E_0\left[E_0\left[\frac{1}{\sigma_t^2}[1, |\varepsilon_{t-1}|, \ldots, |\sigma_{t-p}|]'/\underline{\varepsilon_{t-2}}\right] \times \sum_{i=2}^{q} \alpha_i (I_{\varepsilon_{t-i}>0} - I_{\varepsilon_{t-i}<0})(-Y_{t-i-1})\right],$$

using the fact that $(1/\sigma_t^2)[1, |\varepsilon_{t-1}|, \ldots, \sigma_{t-p}]'\alpha_1(I_{\varepsilon_{t-1}>0} - I_{\varepsilon_{t-1}<0})(-Y_{t-2})$ is an odd function in ε_{t-1} and that this variable has a symmetric conditional distribution. Now $E_0[(1/\sigma_t^2)[1, |\varepsilon_{t-1}|, \ldots, |\varepsilon_{t-q}|]'/\underline{\varepsilon_{t-2}}]$ is an even function of ε_{t-2} and the same method applies. Iterating the procedure proves the property. Q.E.D.

References

Abramowitz, M. and I.A. Stegun, 1964, Handbook of mathematical functions (Dover, New York, NY).

Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman, 1974, Estimation inference in nonlinear structural models, Annals of Economic and Social Measurement 4, 653–665.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327.

Christie, A.A., 1982, The stochastic behavior of common stock variances, Journal of Financial Economics 10, 407–432.

Davidian, M. and R.J. Carroll, 1987, Variance function estimation, Journal of the American Statistical Association 82, 1079–1091.

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987–1008.

Engle, R.F., D. Lilien, and R. Robins, 1985, Estimation of time varying risk premiums in the term structure, Econometrica 55, 391–408.

Glosten, L.R., R. Jagannathan, and D. Runkle, 1989, Relationship between the expected value and the volatility of the nominal excess return on stocks, Discussion paper (Northwestern University, Evanston, IL).

Godfrey, L.J., 1981, On the invariance of the Lagrange multiplier test with respect to certain changes in the alternative hypothesis, Econometrica 49, 1443–1455.

Gouriéroux, C. and A. Monfort, 1989, Statistiques et modèles econometriques (Economica, Paris).

Gouriéroux, C., 1989, Modèles ARCH et applications à la finance, Discussion paper (Economica, Paris).

Huber, J., 1967, The behavior of maximum likelihood estimates under nonstandard conditions, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1 (University of California, Berkeley, CA).

Milhoj, A., 1984, The moment structure of ARCH processes, Scandinavian Journal of Statistics, 12, 281–292.

Nelson, D.B., 1990a, Stationarity and persistence in the GARCH(1, 1) model, Econometric Theory, 6, 318–334.

Nelson, D.B., 1990b, A note on the normalized residuals from ARCH and stochastic volatility models, Discussion paper (University of Chicago, Chicago, IL).

Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59, 347–370.

Nelson, D.B. and C.Q. Cao, 1992, A note on the inequality constraints in the univariate GARCH model, Discussion paper (University of Chicago, Chicago, IL).

Schwert, G.W., 1989, Why does stock market volatility change over time?, Journal of Finance XLIV, 1115–1153.

Sentana, E., 1990, Quadratic ARCH models: A potential re-interpretation of ARCH models as second order Taylor approximations, Discussion paper (London School of Economics, London).

Taylor, S., 1986, Modelling financial time series (Wiley, New York, NY).

Tong, H., 1990, Nonlinear time series: A dynamical system approach (Clarendon Press, Oxford).

Weiss, A.A., 1984, ARMA models with ARCH errors, Journal of Time Series Analysis 5, 129–143.

Weiss, A.A., 1986, Asymptotic theory for ARCH models: Estimating and testing, Econometric Theory 2, 107–131.