Contemporaneous asymmetry in GARCH processes

Mohamed El Babsiri\textsuperscript{a}, Jean-Michel Zakoian\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a}CDC-marchés, 3 rue Lafayette, 75009 Paris, France
\textsuperscript{b}GREMARS, Université de Lille 3, CREST (Centre de Recherche en Economie et Statistique), 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France

Received 29 July 1997; received in revised form 3 August 2000; accepted 2 October 2000

Abstract

The paper introduces a new concept of asymmetry (contemporaneous asymmetry) in conditional heteroskedasticity models. We propose an original class of models aimed to capture the leverage effect, contemporaneous asymmetry as well as time-varying skewness and kurtosis. Not only past up and down moves have different impacts on the conditional variance, but also, positive and negative changes are governed by different conditional variances. We give conditions for the existence of a second-order and strictly stationary solution. The paper also provides consistency results on the quasi-maximum likelihood estimation. Finally, an empirical analysis on the French CAC 40 stock index is proposed. © 2001 Elsevier Science S.A. All rights reserved.

\textit{JEL classification:} C13; C32

\textit{Keywords:} Contemporaneous asymmetry; Conditional kurtosis; GARCH; Stationarity; Quasi-maximum-likelihood

1. Introduction

In the empirical finance literature, it has been recognized for a long time that stock return volatility is changing over time. For example, French et al. (1987)
and Schwert (1989) have shown that estimates of the standard deviation of monthly stock returns can fluctuate dramatically over time. Similar results hold for daily volatilities (see, e.g., Schwert, 1990). Other important characteristics of financial series are the heavy-tailed distributions and the clusters of outliers (large changes in stock returns are more common than would be expected with a normal distribution). Finally, one of the best documented stylized facts characterizing stock market returns is that there is a tendency for changes in volatility to be negatively correlated with changes in stock prices. This dynamic asymmetry, or ‘leverage effect’, was first noted by Black (1976) and confirmed by many authors, at least for short-run impacts (see, e.g., Christie, 1982; Nelson, 1990; Campbell and Hentschel, 1992; Engle and Ng, 1993).

A number of econometric models have attempted to account for these empirical findings. Among them the family of GARCH models, introduced by Engle (1982) and generalized by Bollerslev (1986), is one of the most fruitful. For an extensive literature we refer to Bollerslev et al. (1994), Gourieroux (1997). In these models, the conditional variance \( \sigma_t^2 \) of the time-\( t \) prediction error \( \varepsilon_t \) is determined endogeneously as a function of lagged variables. The models of the first generation, based on squares as in the original paper, or absolute values (Taylor, 1986; Schwert, 1989), did not capture dynamic asymmetry. In particular, they did not allow for correlation between returns and future volatility. To overcome this limitation, a second generation of GARCH-type models was introduced by Nelson (1991), Higgins and Bera (1992), Glosten et al. (1993), Zakoian (1994), Sentana (1995) among others.

Most of the univariate GARCH specifications encountered in the literature can be nested in the class proposed by Ding et al. (1995), in which the volatility dynamics is given by

\[
[I - B(L)]\sigma_t^2 = \omega + A_+ (L) (\varepsilon_t^+)^2 + A_- (L) (-\varepsilon_t^-)^2, \quad \lambda > 0,
\]

where \( \varepsilon_t^+ = \max(\varepsilon_t, 0) \) and \( \varepsilon_t^- = \min(\varepsilon_t, 0) \) are the positive and negative parts of the innovation and \( B(L), A_+ (L), A_- (L) \) are finite-order lag polynomials in the lag operator \( L \), which are assumed to fulfill some invertibility and nonnegativity conditions. Symmetric GARCH correspond to \( A_+ (\cdot) = A_- (\cdot) \). Recently, following Hentschel (1995), a generalization has been proposed by Liu et al. (1997) based on the juxtaposition of a piecewise linear conditional mean and a piecewise linear conditional variance. The exponential GARCH of Nelson (1991), which does not belong to the previous class due to a multiplicative modeling of volatility as a function of past rescaled residuals \( Z_t = \sigma_t^{-1} \varepsilon_t \) (instead of innovations), also successfully accounts for asymmetries.

An important limitation of the commonly used GARCH models, is that they rely on strong or semi-strong (see Drost and Nijman, 1993; Nijman and Sentana, 1996) distributional assumptions on the error process. Typically, the sequence of innovations driving the model is assumed to be a martingale difference. Frequently, it is even assumed that the rescaled innovations \( \sigma_t^{-1} \varepsilon_t \) are...
conditionally i.i.d. \((0, 1)\) or, further, that they are \(\mathcal{N}(0, 1)\). A crucial consequence of the martingale difference assumption concerns the modeling of asymmetry. Although asymmetric GARCH models allow positive and negative changes to have different impacts on future volatilities, the two components of the innovation have – up to a constant – the same volatilities. To see why this is unsatisfactory, suppose at time \(t\), past negative innovations have been typically of higher magnitudes than positive ones. Then an unanticipated decrease is likely to be larger than an unanticipated increase. So it is desirable to allow for an asymmetric confidence interval around the prediction value.

There are, in fact, economic reasons for believing that the constraints implied by GARCH-type models are too restrictive in the finance context. The set of past returns at a given time is informative about the volatility of good news (‘positive’ risk) and the volatility of bad news (‘negative’ risk). Making their investments, economic agents should account for these positive and negative risks since a seller, for example, is likely to be positive risk-averse but not necessarily negative risk-averse. In the finance literature, the problem of optimal portfolio selection under limited downside risk has been studied for a long time (see, e.g., Arzac and Bawa, 1977; Bawa, 1978; Yamaguchi, 1994; Jansen et al., 1998). Hence, from an economic point of view, it makes sense to look for a model able to approximate the volatilities of bad and good news.

Another obvious consequence of strong independence assumptions on the rescaled innovations \(\sigma_t^{-1} \varepsilon_t\) (e.g., independence of the \(Z_t\)’s) is that all conditional moments of \(\varepsilon_t\) are proportional, with a time-independent proportionality constant. In particular, the conditional kurtosis and skewness of \(\varepsilon_t\) are that of \(Z_t\), i.e., they are fixed. However, it is easy to check on financial series that these quantities can have huge fluctuations over time, just as conditional variances. Within the GARCH framework, one solution to the conditional kurtosis problem has been proposed by Hansen (1994). In the context of continuous time pricing models, Das and Sundaram (1998) have introduced a structure aimed to generate a term structure of skewness and kurtosis.

The aim of the present paper is to develop a structure, allowing for time-varying skewness and kurtosis (heteroskewness and heterokurtosis) and two kinds of asymmetry: (i) different volatility processes for up and down moves in equity markets (contemporaneous asymmetry); (ii) asymmetric reactions of these volatilities to past positive and negative changes (dynamic asymmetry or leverage effect). The paper is organized as follows. The concept of contemporaneous asymmetry is presented in Section 2. The general model is laid out in Section 3. Due to its local linearity property, the advanced class of models remains very tractable and we are able to establish sufficient conditions for second-order stationarity. This is done in Section 4. Section 5 is devoted to the consistency of two quasi-maximum likelihood estimators. Section 6 proposes an empirical study based on the French stock index. Another objective of the paper is to show
that contemporaneous asymmetry is significantly present in the data, even more pronounced than the leverage effect. We also compare the empirical performance of our model with other stochastic variance specifications. Section 7 is a conclusion. All proofs are reported in the appendices.

Throughout the paper, the positive and negative parts of any real random variable $X$ are noted as $X^+ = \max(X, 0)$ and $X^- = \min(X, 0)$, respectively. The space of square-integrable real processes is noted as $L^2$.

2. Contemporaneous asymmetry

For any stationary process $\{\varepsilon_t\}$ belonging to $L^2$, it will be fruitful to consider the elementary decomposition $\varepsilon_t = \varepsilon_t^+ + \varepsilon_t^-$. Then, let $\sigma_{t,+} = E(\varepsilon_t^+ | \varepsilon_{t-1})$ and $\sigma_{t,-} = E(-\varepsilon_t^- | \varepsilon_{t-1})$, where $\varepsilon_{t-1}$ denotes the past of $\varepsilon_t$, i.e., the $\sigma$-field generated by $\{\varepsilon_s; s < t\}$. Assuming that the positive (resp. negative) part of $\varepsilon_t$ is non-degenerate, we have almost surely (a.s.) $\sigma_{t,+} > 0$ (resp. $\sigma_{t,-} > 0$). Therefore, we can set $Z_{t,+} := \sigma_{t,+}^{-1}\varepsilon_t^+$ and $Z_{t,-} := \sigma_{t,-}^{-1}\varepsilon_t^-$. Let $Z_t := Z_{t,+} + Z_{t,-}$. Since $Z_{t,+} > 0$, $Z_{t,-} < 0$ and $Z_{t,+}Z_{t,-} = 0$, it is clear that $Z_{t,+}$ and $Z_{t,-}$ are the positive and negative parts of a process $(Z_t)$: $Z_{t,+} = Z_t^+$, $Z_{t,-} = Z_t^-$. Finally, the following representation holds:

$$\varepsilon_t = \sigma_{t,+}Z_t^+ + \sigma_{t,-}Z_t^-,$$

where $\sigma_{t,+}$ and $\sigma_{t,-}$ are positive variables belonging to the past of $\varepsilon_t$, $\{Z_t\}$ is a process such that $E(Z_t^+ | \varepsilon_{t-1}) = E(Z_t^- | \varepsilon_{t-1}) = 1$. The representation is obviously unique. Now we have

$$E(\varepsilon_t | \varepsilon_{t-1}) = \sigma_{t,+} - \sigma_{t,-}.$$

**Definition 1.** We call $\{\varepsilon_t\}$ contemporaneously asymmetric if the random variables $\varepsilon_t^+$ and $-\varepsilon_t^-$ have different conditional expectations (i.e., if $\sigma_{t,+} - \sigma_{t,-}$ is not (a.s.) equal to zero).

It is clear that no semi-strong GARCH process can meet the contemporaneous asymmetry requirement since when the martingale difference assumption holds, we have $\sigma_{t,+} = \sigma_{t,-} := \sigma_t$. In such a case, we simply have $\varepsilon_t = \sigma_tZ_t$ and we call $\{\varepsilon_t\}$ contemporaneously symmetric.

A current assumption in the GARCH framework is that $\{Z_t\}$ is an i.i.d. process, with $Z_t$ independent of the past of $\varepsilon_t$. We examine some straightforward consequences of the definition under this assumption.
2.1. Conditional distribution

A first step in the direction of modeling the conditional distribution was taken by Hansen (1994), who directly specified it as a function of the information set, depending on a low-dimensional parameter vector. If contemporaneous asymmetry holds, the conditional distributions of $e_t^+$ and $-e_t^-$ are different. The converse is not necessarily true: a simple example is obtained by taking a semi-strong GARCH with an asymmetric distribution for $Z_t$. However, in the classical GARCH framework $Z_t$ and $e_t$ have similar conditional density functions (e.g., normal, unimodal, symmetric, asymmetric, etc.). In the present framework, the conditional distribution of $e_t$ is a mixture of the distributions of $Z_t^+$ and $Z_t^-$, with time-varying mixing coefficients. Assume, for instance, that the density function of $Z_t$ is unimodal. Then, graphically, the conditional distribution of $e_t$ for the case of $\sigma_{t,+} < \sigma_{t,-}$ will be as in Fig. 1. At time $t + 1$ the shape of the conditional distribution will be different.

When the variables $\sigma_{t,+}$ and $\sigma_{t,-}$ are nondegenerate, the positive and negative parts of $e_t$ are of course conditionally heteroskedastic, with conditional variances given by $\sigma_{t,+}^2 V(Z_t^+)$ and $\sigma_{t,-}^2 V(Z_t^-)$, respectively. Moreover,

$$V(e_t|e_{t-1}) = \sigma_{t,+}^2 V(Z_t^+) + \sigma_{t,-}^2 V(Z_t^-) + 2\sigma_{t,+}\sigma_{t,-}. \tag{2}$$

2.2. Heteroskewness, heterokurtosis

In the semi-strong GARCH setting, the conditional skewness and kurtosis of $e_t$ are time-independent and equal to the skewness and kurtosis of $Z_t$. 

Fig. 1. Conditional density function of $e_t$. 

A first step in the direction of modeling the conditional distribution was taken by Hansen (1994), who directly specified it as a function of the information set, depending on a low-dimensional parameter vector. If contemporaneous asymmetry holds, the conditional distributions of $e_t^+$ and $-e_t^-$ are different. The converse is not necessarily true: a simple example is obtained by taking a semi-strong GARCH with an asymmetric distribution for $Z_t$. However, in the classical GARCH framework $Z_t$ and $e_t$ have similar conditional density functions (e.g., normal, unimodal, symmetric, asymmetric, etc.). In the present framework, the conditional distribution of $e_t$ is a mixture of the distributions of $Z_t^+$ and $Z_t^-$, with time-varying mixing coefficients. Assume, for instance, that the density function of $Z_t$ is unimodal. Then, graphically, the conditional distribution of $e_t$ for the case of $\sigma_{t,+} < \sigma_{t,-}$ will be as in Fig. 1. At time $t + 1$ the shape of the conditional distribution will be different.

When the variables $\sigma_{t,+}$ and $\sigma_{t,-}$ are nondegenerate, the positive and negative parts of $e_t$ are of course conditionally heteroskedastic, with conditional variances given by $\sigma_{t,+}^2 V(Z_t^+)$ and $\sigma_{t,-}^2 V(Z_t^-)$, respectively. Moreover,

$$V(e_t|e_{t-1}) = \sigma_{t,+}^2 V(Z_t^+) + \sigma_{t,-}^2 V(Z_t^-) + 2\sigma_{t,+}\sigma_{t,-}. \tag{2}$$

2.2. Heteroskewness, heterokurtosis

In the semi-strong GARCH setting, the conditional skewness and kurtosis of $e_t$ are time-independent and equal to the skewness and kurtosis of $Z_t$. 

A first step in the direction of modeling the conditional distribution was taken by Hansen (1994), who directly specified it as a function of the information set, depending on a low-dimensional parameter vector. If contemporaneous asymmetry holds, the conditional distributions of $e_t^+$ and $-e_t^-$ are different. The converse is not necessarily true: a simple example is obtained by taking a semi-strong GARCH with an asymmetric distribution for $Z_t$. However, in the classical GARCH framework $Z_t$ and $e_t$ have similar conditional density functions (e.g., normal, unimodal, symmetric, asymmetric, etc.). In the present framework, the conditional distribution of $e_t$ is a mixture of the distributions of $Z_t^+$ and $Z_t^-$, with time-varying mixing coefficients. Assume, for instance, that the density function of $Z_t$ is unimodal. Then, graphically, the conditional distribution of $e_t$ for the case of $\sigma_{t,+} < \sigma_{t,-}$ will be as in Fig. 1. At time $t + 1$ the shape of the conditional distribution will be different.

When the variables $\sigma_{t,+}$ and $\sigma_{t,-}$ are nondegenerate, the positive and negative parts of $e_t$ are of course conditionally heteroskedastic, with conditional variances given by $\sigma_{t,+}^2 V(Z_t^+)$ and $\sigma_{t,-}^2 V(Z_t^-)$, respectively. Moreover,

$$V(e_t|e_{t-1}) = \sigma_{t,+}^2 V(Z_t^+) + \sigma_{t,-}^2 V(Z_t^-) + 2\sigma_{t,+}\sigma_{t,-}. \tag{2}$$
Contrariwise, when contemporaneous asymmetry holds, the conditional skewness and kurtosis are time-varying. To quantify this, let us introduce the strong innovation process \( \{ \eta_t \} \) of \( \{ \epsilon_t \} \), defined as \( \eta_t = \epsilon_t - E(\epsilon_t | \epsilon_{t-1}) = \sigma_{t,+} U_{t,+} + \sigma_{t,-} U_{t,-} \), for all \( t \in \mathbb{Z} \), where \( U_{t,+} = Z^+_t - E(Z^+_t) \) and \( U_{t,-} = Z^-_t - E(Z^-_t) \) are i.i.d. centered processes, independent of the past. Let \( c(k,l) = E[U_{t,+} U_{t,-}] \), \( k, l \in \mathbb{N} \). From these definitions, the conditional skewness and kurtosis of \( \{ \epsilon_t \} \) are easily shown to be, respectively, equal to

\[
\begin{align*}
S_t &= \frac{E[\eta_t^3 | \epsilon_{t-1}]}{[E[\eta_t^2 | \epsilon_{t-1}]]^{3/2}} = \sum_{k=0}^{3} \binom{3}{k} \sigma_{t,+}^k \sigma_{t,-}^{3-k} c(k, 3-k) \\
K_t &= \frac{E[\eta_t^4 | \epsilon_{t-1}]}{[E[\eta_t^2 | \epsilon_{t-1}]]^2} = \sum_{k=0}^{4} \binom{4}{k} \sigma_{t,+}^k \sigma_{t,-}^{4-k} c(k, 4-k)
\end{align*}
\]  

From these expressions, \( S_t \) and \( K_t \) can be expressed as functions of the ratio \( \sigma_{t,+} / \sigma_{t,-} \). When it is equal to one at date \( t \), the conditional skewness and kurtosis coincide with those of the i.i.d. process \( \{ Z_t \} \). This is in particular the case, at all dates, when contemporaneous asymmetry does not hold \( (\sigma_{t,+} = \sigma_{t,-}) \). From an empirical point of view, these calculations show that the concept of contemporaneous asymmetry should allow to reconsider the problem of negative skewness or excess kurtosis arising in financial data.

2.3. Autocorrelation structure of \( \{ \epsilon_t^+, \epsilon_t^- \} \)

In this section, we characterize both types of symmetries in terms of the autocorrelation structure of the vector \( \{ \epsilon_t^+, \epsilon_t^- \} \).

Contemporaneous symmetry implies that \( \text{cov}(\epsilon_t, z_{t-1}) = 0 \), for all variable \( z_{t-1} \in \epsilon_{t-1}, z_{t-1} \in L^2 \). In particular, we have

\[ \forall k > 0, \quad \text{cov}(\epsilon_t, \epsilon_{t-k}^+) = \text{cov}(\epsilon_t, \epsilon_{t-k}^-) = 0. \]

Similarly, we can characterize dynamic symmetry (i.e., absence of the so-called leverage effect) by

\[ \forall k > 0, \quad \text{cov}(\epsilon_t^+, \epsilon_{t-k}^-) = \text{cov}(\epsilon_t^-, \epsilon_{t-k}^-) = 0. \]

In the usual GARCH framework \( (\epsilon_t = \sigma_t Z_t) \), the last condition is equivalent to:

\( (\forall k > 0) \text{cov}(\sigma_t, \epsilon_{t-k}) = 0 \), which can also be written as \( \text{cov}(\sigma_t, \epsilon_{t-k}^+) = \text{cov}(\sigma_t, - \epsilon_{t-k}^-) \). Conversely, dynamic asymmetry means that the past positive
and negative values of $e_t$ (for at least one lag $k$), have different impacts on the current volatility.

Let us remark that when (at least) one form of symmetry holds, $\{e_t\}$ has automatically the autocovariance structure of a white noise. When both types of asymmetries are ruled out, $\{e_t^+, e_t^-\}$ is nonautocorrelated.

3. Model specification

In this section, we propose a class of GARCH-type models able to capture the features presented in Section 2. Therefore, we have to parameterize $\{\sigma_{t, +}\}$ and $\{\sigma_{t, -}\}$ as positive functions of the $e_t$ past values. To accommodate contemporaneous asymmetry we need different specifications for these variables. Moreover, to allow for dynamic asymmetry we have to parameterize $\{\sigma_{t, +}\}$ and $\{\sigma_{t, -}\}$ as asymmetric functions of the past positive and negative innovations. A natural choice, which makes the model very tractable, is to choose them linear in these past positive and negative parts of the noise.

We therefore consider the following model:

$$e_t = \sigma_{t, +} Z_t^+ + \sigma_{t, -} Z_t^-, \quad t \in \mathbb{Z},$$

where $\{Z_t^+\}$ is an i.i.d. centered process, $Z_t$ is independent of $e_{t-1}$, $Z_t \in L^2$ for all $t \in \mathbb{Z}$, and

$$\sigma_{t, +} = \alpha_{0, +} + \sum_{i=1}^q \alpha_{i, +} e_{t-i}^+ - \alpha_{i, -} e_{t-i}^- + \sum_{j=1}^p \beta_{j, +} \sigma_{t-j, +} + \beta_{j, -} \sigma_{t-j, -},$$

$$\sigma_{t, -} = \alpha_{0, -} + \sum_{i=1}^q \alpha_{i, +} e_{t-i}^- - \alpha_{i, -} e_{t-i}^+ + \sum_{j=1}^p \beta_{j, -} \sigma_{t-j, +} + \beta_{j, +} \sigma_{t-j, -},$$

where $\{\alpha_{i, +}\}, \{\alpha_{i, -}\}, \ldots, \{\beta_{j, -}\}$ are real, nonstochastic scalar sequences; $\alpha_{0, +}$, $\alpha_{0, -}$ are real numbers.

Without loss of generality, we can assume that $E(Z_t^+) = E(-Z_t^-) = 1$.

The model can be written in the equivalent form as

$$[I - B(L)] \begin{bmatrix} \sigma_{t, +} \\ \sigma_{t, -} \end{bmatrix} = \alpha_0 + A(L) \begin{bmatrix} e_t^+ \\ -e_t^- \end{bmatrix},$$

where $B(L) = \sum_{j=1}^p B_j L^j$ and $A(L) = \sum_{j=1}^q A_j L^j$ are finite-order lag matrix polynomials in the lag operator $L$; $\alpha_0 = (\alpha_{0, +}, \alpha_{0, -})$.

We make the following assumptions on the lag polynomials:

(A1) $I - B(L)$ is invertible (i.e., the roots of its determinant lie outside the unit circle);

(A2) $[I - B(L)]$ and $A(L)$ have no common roots and $[I - B(L)]^{-1} A(L)$ (resp. $[I - B(L)]^{-1} \alpha_0$) has nonnegative (resp. positive) coefficients.
From (A1), $\sigma_{t,+}$ and $\sigma_{t,-}$ can be expressed as linear functions of the past values of $e_t^+$ and $e_t^-$. From (A2), $\sigma_{t,+} > 0$ and $\sigma_{t,-} > 0$. Of course, in empirical applications, it is simplest (although not equivalent, see Nelson and Cao (1992)) to assume the nonnegativity of each coefficient in (5) instead of (A2). A crucial consequence of the positivity of $\sigma_{t,+}$ and $\sigma_{t,-}$ is that

$$
e_t^+ = \sigma_{t,+} Z_t^+ \quad \text{and} \quad \ne_t^- = \sigma_{t,-} Z_t^-.$$  

(7)

This will make possible the analysis of the probability structure of the model. Another important consequence of (7) is to allow for interpretation: $\sigma_{t,+}$ and $\sigma_{t,-}$ are, up to constants, the conditional standard deviations of $e_t^+$ and $e_t^-$, respectively. As a particular case, the class contains the GTARCH model introduced by Zakoïan (1994) in which only dynamic asymmetry holds (by imposing equal coefficients in the dynamics of $\sigma_{t,+}$ and $\sigma_{t,-}$). It should be noted that, apart from the GTARCH case, dynamic asymmetry cannot be straightforwardly characterized in terms of the model coefficients.

It is important to notice that the probabilities of occurrence of positive and negative shocks are independent of the past (since the sign of $e_t$ is that of $Z_t$). Moreover, if we assume that the median of the distribution of $\{Z_t\}$ is zero, the probabilities are equal. However, conditionally to the past, the expected magnitudes of a positive and a negative shock are different. Finally, these expected magnitudes are a function of both the magnitudes and the signs of past positive shocks.

While the functional form of the model is rather complicated it can be useful to focus on the effects of news on volatility. A paper by Engle and Ng (1993) put forth the idea of a news impact curve which characterizes the impact of past return shocks on the return volatility. Holding constant the information dated $t - 2$ and earlier, it consists in examining the implied relation between $e_{t-1}$ and the time-$t$ volatility. Our model differs from the standard GARCH-type models in that it allows for different functional form of the volatilities of positive and negative innovations. Therefore, we will consider the two corresponding news impact curves. In general, both of them have different slopes for their positive and negative parts. In the following example (Fig. 2), we assume that the volatilities of the positive and negative parts of $e_t$ display inverted asymmetries. It is seen that an unexpected drop in price at time $t - 1$ ($e_{t-1} < 0$) increases the volatility ($\sigma_{t,-}$) of a decrease at time $t$ more than an unexpected increase in price ($e_{t-1} > 0$) of similar magnitude. Conversely, as far as the volatility $\sigma_{t,+}$ of an increase at time $t$ is concerned, an unexpected drop decreases volatility while an unexpected increase has a positive impact. In each case, the impact on volatility is proportional to the magnitude of the shock. Depending on the model specification, the slopes of the news impact curves in the positive and negative sides can of course be very different from those presented here.
The news impact curve of \( \sigma_t := \sqrt{\text{Var}(\epsilon_t|\epsilon_{t-1})} \) is also plotted in the same figure. It should be noted that, the two sides of the curves are nonlinear. From Eq. (2), it is seen that this nonlinearity does not occur when \( V(Z_t^+) = V(Z_t^-) = 1 \) in which case \( \sigma_t = \sigma_{t,+} + \sigma_{t,-} \).

4. Stationarity

In this section, we derive sufficient conditions for the existence of weak stationary solutions to model (4), (5), in the lines of Bougerol and Picard (1992).\(^1\) We also give sufficient conditions for the existence of white noise solutions.

\(^1\) Necessary and sufficient conditions for wide-sense and strict stationarity of threshold GARCH models of the form (1) (with \( \lambda = 1 \)) have been established by Zakoian (1994) in the GARCH(1,1) case. They have been generalized to higher-order lag polynomials by Gonçalves and Mendes Lopes (1994, 1996).
We need to introduce the following notations. For convenience we can always suppose that \( p > 1, q > 2 \), by adding some coefficients equal to zero if needed. For all \( t \in \mathbb{Z} \), let

\[
X_t = (\sigma_{t+1,+}, \sigma_{t+1,-}, \ldots, \sigma_{t-p+2,+}, \sigma_{t-p+2,-}, e^+_t, \\
- \varepsilon_t, \ldots, e_{t-q+2}, - \varepsilon_{t-q+2})' \in \mathbb{R}^{2(p+q-1)}
\]

and

\[
\begin{align*}
\alpha_+ &= (\alpha^+_2, \alpha^-_2, \ldots, \alpha^-_{q-1}, +) \in \mathbb{R}^{2(q-2)}, \\
\alpha_- &= (\alpha^+_2, \alpha^-_2, \ldots, \alpha^-_{q-1}, -) \in \mathbb{R}^{2(q-2)}, \\
a_+(Z_t) &= (\beta^+_1, + + \alpha^+_1, + Z^+_1, \beta^-_1, + - \alpha^-_1, + Z^-_1, \\
&\quad \beta^+_2, +, \beta^-_2, +, \ldots, \beta^+_p, +, \beta^-_p, +) \in \mathbb{R}^{2(p-1)}, \\
a_-(Z_t) &= (\beta^+_1, - + \alpha^+_1, - Z^+_1, \beta^-_1, - - \alpha^-_1, - Z^-_1, \\
&\quad \beta^+_2, -, \beta^-_2, -, \ldots, \beta^+_p, -, \beta^-_p, -) \in \mathbb{R}^{2(p-1)}, \\
b_+(Z_t) &= (Z^+_t, 0, 0, \ldots, 0) \in \mathbb{R}^{2(p-1)}, \quad b_-(Z_t) = (0, - Z^-_t, 0, \ldots, 0) \in \mathbb{R}^{2(p-1)}.
\end{align*}
\]

Define the \( 2(p+q-1) \times 2(p+q-1) \) matrix \( C(Z_t) \), written in block-form, by

\[
C(Z_t) = \begin{pmatrix}
a_+(Z_t) & \beta^+_p, + & \beta^-_p, + & \alpha^+_p & \alpha^-_p \\
a_-(Z_t) & \beta^+_p, - & \beta^-_p, - & \alpha^+_p & \alpha^-_p \\
I_{2(p-1)} & 0 & 0 & 0 & 0 \\
b_+(Z_t) & 0 & 0 & 0 & 0 \\
b_-(Z_t) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & I_{2(q-2)} & 0 & 0
\end{pmatrix}
\]

and \( b = (\alpha_{0,+}, \alpha_{0,-}, 0, \ldots, 0)' \in \mathbb{R}^{2(p+q-1)} \). Then, an equivalent representation of (6) is

\[
X_{t+1} = C(Z_{t+1})X_t + b.
\] (8)

Since \( \{C(Z_t)\} \) is a sequence of i.i.d. square matrices, with \( C(Z_{t+1}) \) independent of \( X_t \), the multivariate linear stochastic difference equation (8) makes \( \{X_t\} \) a Markov process. This is used to establish the stationarity results. We use \( \rho(M) = \max \{ |\text{all eigenvalues of } M| \} \) to denote the spectral radius of a matrix \( M \). We first give the following sufficient condition for weak stationarity.
Theorem 1. Let $\tilde{C} = E[C(Z_t)]$ and $C^* = E[C(Z_t) \otimes C(Z_t)]$. Then if
\[ \max(\rho(\tilde{C}), \rho(C^*)) < 1, \]
model (4), (5) has a unique second-order stationary solution $\{\tilde{e}_t\}$. Moreover, the solution is also strictly stationary and ergodic.

Once the stationarity conditions are fulfilled, the following result is helpful to obtain the autocovariance structure (till lag $q$) of $\{\tilde{e}_t, \tilde{e}_{t-1}\}$.

Corollary 1. Suppose that the assumptions of Theorem 1 hold. Let $W_t = [\text{Vec}(X_t, X'_t), X'_t], \tilde{b} = [\text{Vec}(bb')']$, and
\[ \tilde{A} = \begin{bmatrix}
C^* & b \otimes \tilde{C} + \tilde{C} \otimes b \\
0 & \tilde{C}
\end{bmatrix}. \]
Then we have $E[W_t] = (I - \tilde{A})^{-1}\tilde{b}$.

Remark. (a) We have obtained conditions ensuring existence of a solution in $L^2$ and it was proved that the solution is also strictly stationary. Under less restrictive conditions, the existence of a strictly stationary solution which is not necessarily squared integrable can be established. More precisely, a necessary and sufficient condition for the existence of a (unique) strictly stationary solution is that the top-Lyapunov exponent associated with the matrices $\{C(Z_t)\}$ in (8) is strictly negative (see El Babsiri and Zakoïan, 1997 for more details).

(b) It is worth noting that, contrary to standard GARCH models, the second-order stationary solution to model (4), (5), when existing, is not necessarily a white noise. This is due to the fact that $\{\tilde{e}_t\}$ is not a martingale difference. However, the model can be constrained to produce a white-noise solution. In a previous draft of the paper, two sets of sufficient conditions were provided, both of them corresponding to a finite number of constraints on the coefficients (see El Babsiri and Zakoïan, 1997).

5. Consistency of quasi-maximum-likelihood estimators

Despite the large empirical literature on GARCH-type processes, the theoretical aspects of the estimation procedures are far from being fully investigated. Several papers, generally limited to the GARCH(1,1) case, have analyzed the behavior of the quasi-maximum likelihood estimator (QMLE) with different requirements on the finiteness of the unconditional variance or the assumed innovation density (see, e.g., Weiss, 1986; Bollerslev and Wooldridge, 1992; Lee and Hansen, 1994; Lumsdaine, 1996; Jeantheau, 1998). Recently, Newey and Steigerwald (1997) have shown that, for a large class of conditionally heteroskedastic models, consistency of a non-Gaussian QMLE requires that either:
(a) the conditional mean is identically zero; or (b) both the assumed and the true innovation densities are symmetric around zero. When these conditions do not hold, they also have shown that an additional location parameter is needed to ensure asymptotic identifiability.

In this section, we will establish the consistency of two sequences of QMLE under different assumptions. At first sight it is tempting to use the theory developed by Newey and Steigerwald (1997). However, rewriting (4) in terms of the conditional mean and variance of \( \varepsilon_t \) as

\[
\varepsilon_t = \sigma_{t,+} - \sigma_{t,-} + \sigma_t u_t,
\]

where \( \sigma_t^2 = V(\varepsilon_t|\varepsilon_{t-1}) \) is given by (2), it is seen that the period-\( t \) innovation \( u_t \) is not in general i.i.d. and independent of the past (we have \( u_t = (\sigma_{t,+}/\sigma_t)(Z_t^+ - 1) + (\sigma_{t,-}/\sigma_t)(Z_t^- + 1) \)). Although our model is not nested in the framework of Newey and Steigerwald (1997), we will obtain some similarities as far as the identification issue is concerned.

The proof is not limited to the GARCH(1,1) case. The general setup is the following. Let \( \{Y_t\} \), be a real ergodic, regular, second-order and strictly stationary process. The statistical model is an ARMA\((P,Q)\) model of the form

\[
\sum_{i=0}^{P} \phi_i Y_{t-i} = \sum_{i=0}^{Q} \psi_i \varepsilon_{t-i},
\]

(10)

where \( \phi_0 = 1, \psi_0 = 1 \) and the polynomials \( \Phi(z) = \sum_{i=0}^{P} \phi_i z^i \) and \( \Psi(z) = \sum_{i=0}^{Q} \psi_i z^i \) have only zeros outside the unit disk \( |z| \leq 1 \) and no common roots. In addition, the linear innovation \( \{\varepsilon_t\} \) of \( \{Y_t\} \) is assumed to be the solution of model (4),(5).

The parameter vector of the model defined by (10),(4),(5) is denoted as \( \theta \). Now, let \( Y_1, \ldots, Y_T \) denote \( T \) observations of process \( \{Y_t\} \), obtained for the true value \( \theta_0 \). Let \( \delta \) be a strictly positive constant. The unknown parameter \( \theta_0 \) is assumed to belong to the interior of the compact parameter space \( \Theta_\delta \) such that, \( \forall \theta \in \Theta_\delta \): (i) the polynomials \( \Phi_\theta(z) = 1 + \sum_{i=1}^{P} \theta_i z^i \) and \( \Psi_\theta(z) = 1 + \sum_{i=1}^{Q} \psi_i z^i \), (resp. \( \det(A_\theta(z)) \) and \( \det(I - B_\theta(z)) \), see Eq. (11)) have only zeros outside the disk \( |z| < 1 + \delta \) and no common roots; (ii) \( \min(\theta_{P+Q+1}, \theta_{P+Q+2}) \geq \delta \) (see (11)); (iii) the polynomials \( (I - B_\theta)^{-1} \) and \( A_\theta \) have nonnegative coefficients. For all \( \theta \in \Theta_\delta \), we consider the sequence \( \{\varepsilon_t(\theta), \sigma_{t,+}(\theta), \sigma_{t,-}(\theta)\} \) uniquely defined by the system

\[
\begin{align*}
\varepsilon_t(\theta) &= [\Psi_\theta(L)]^{-1}\Phi_\theta(L)Y_t, \\
\begin{bmatrix} \sigma_{t,+}(\theta) \\ \sigma_{t,-}(\theta) \end{bmatrix} &= [I - B_\theta(L)]^{-1}\begin{bmatrix} \theta_{P+Q+1} \\ \theta_{P+Q+2} \end{bmatrix} + A_\theta(L)\begin{bmatrix} \varepsilon_t^+(\theta) \\ -\varepsilon_t^-(\theta) \end{bmatrix}.
\end{align*}
\]

(11)

Note that \( (\varepsilon_t(\theta_0), \sigma_{t,+}(\theta_0), \sigma_{t,-}(\theta_0)) = (\varepsilon_t, \sigma_{t,+}, \sigma_{t,-}) \), a.s., \( \forall t \in \mathbb{Z} \).
The variables $e_t(\theta)$, $\sigma_t^+(\theta)$ and $\sigma_t^-(\theta)$ can be approximated, for $t = 1, \ldots, T$, by $e_t(\theta)$, $s_t^+(\theta)$ and $s_t^-(\theta)$, respectively, obtained using the same recursive formulas (11) but setting all unknown starting values to zero.

The Gaussian quasi-likelihood function (i.e., corresponding to the assumption that $Z_t$ is Gaussian $(0,1)$, although we do not need to make this assumption) for the sample of $T$ observations is

$$L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ell_t(\theta),$$

where (ignoring constants)

$$\ell_t(\theta) = \log \left( \frac{1}{s_{t,+}(\theta)} \exp \left( - \frac{e_t^2(\theta)}{2s_{t,+}^2(\theta)} \right) \mathbf{1}_{e_t(\theta) > 0} \right)$$

$$+ \frac{1}{s_{t,-}(\theta)} \exp \left( - \frac{e_t^2(\theta)}{2s_{t,-}^2(\theta)} \right) \mathbf{1}_{e_t(\theta) < 0}$$

$$= - \left( \log s_{t,+}(\theta) + \frac{e_t^2(\theta)}{2s_{t,+}^2(\theta)} \right) \mathbf{1}_{e_t(\theta) > 0}$$

$$- \left( \log s_{t,-}(\theta) + \frac{e_t^2(\theta)}{2s_{t,-}^2(\theta)} \right) \mathbf{1}_{e_t(\theta) < 0}.$$
assumption. Therefore, we set
\[ \hat{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{z}_t(\theta), \]  
where
\[ \tilde{z}_t(\theta) = -\left( \log s_t(\theta) + \frac{(e_t(\theta) - s_{t,+}(\theta) + s_{t,-}(\theta))^2}{2s_t^2(\theta)} \right) \]
where
\[ s_t(\theta)^2 = s_{t,+}^2(\theta)V(Z_t^+) + s_{t,-}^2(\theta)V(Z_t^-) + 2s_{t,+}(\theta)s_{t,-}(\theta). \]

Let \( \hat{\theta}_T \) denote the modified QML estimator obtained by maximizing \( \hat{L}_T(\theta) \). The next result shows that \( \hat{\theta}_T \) is consistent under additional moment assumptions on the i.i.d. process \( \{Z_t\} \). Here we need to be more specific concerning the first two moments of the vector \( \{Z_t^+, Z_t^-\} \).

Theorem 3. Suppose that the variances of \( Z_t^+ \) and \( Z_t^- \) are known. Then,
\[ \hat{\theta}_T \to \theta_0 \quad \text{a.s. as } T \to \infty. \]

From the proof of the theorem, the assumption that the variances of the positive and negative parts of \( (Z_t) \) equal to 1, is necessary to ensure that \( \theta_0 \) is the unique maximum of the asymptotic criterion. Otherwise, \( \theta_0 \) is indeed a maximum but uniqueness is not guaranteed.

Once consistency of the estimator has been proved, the next step is obviously to establish sufficient conditions for its asymptotic normality. However, this is substantially more difficult and left for future work. Indeed, the behaviour of the loglikelihood is far from being standard since it is in general not even continuous in the conditional mean parameters (although there is only a countable set of discontinuity points). In the case of lack of contemporaneous asymmetry (condition (i) in Theorem 2), the loglikelihood is continuous but remains nondifferentiable for certain values of the conditional mean parameters (such that \( e_t = 0 \)).\(^2\)

Therefore, some elaborate techniques for proving asymptotic normality, such as Huber’s (1967), seem to be required.

6. Empirical application to the French stock index

Black’s (1976) empirical observation that stock volatility tends to rise following negative returns and to drop following positive ones has motivated a large

\(^2\) However, because the measure of the parameter set where the nondifferentiability occurs is equal to zero, the problem does not arise in practical applications.
number of empirical studies based on GARCH-type models. For instance, Glosten et al. (1993), Rabemananjara and Zakoïan (1993), Zakoïan (1994) found evidence of asymmetries in the volatility of American and French stock returns. However, for the latter returns, Black’s intuition was far from being fully supported by the data. It was shown that: (i) the relative impacts of negative and positive shocks on the current volatility are not the same at all lags; (ii) small shocks and large shocks can provide inverted asymmetries. Our aim in this section is to find empirical evidence for the introduction of different dynamics for positive and negative shocks.

The data consist of a series of a daily stock index: the French CAC 40, which is a value-weighted index computed from 40 stocks traded on the monthly Settlement Market. It is built to represent the whole range of equities listed on the Paris Bourse and it is used in futures and options markets. The data cover the period from 3 November 1988 to 3 November 1998 involving 2485 observations. However, we used only 2385 observations to estimate the models, keeping 100 observations for out-of-sample comparisons.

Denote $S_t$ as the price index and $r_t = \log S_t - \log S_{t-1}$ as the compounded return at time $t$. In Table 1, some summary statistics for $r_t$ are reported. We can see that the series is negatively skewed, so large negative returns are more common than large positive ones. Moreover, the excess kurtosis for $r_t$ indicates the existence of a characteristic fat-tailed behaviour compared with a normal distribution. The Jarque and Bera (1980) normality test statistic is far beyond the critical value which suggests that $r_t$ is far from a normal distribution.

Before the general model is fitted to the data, its underlying assumptions have to be verified empirically. The main point is the occurrence of asymmetries in the conditional distributions. Recall that, from Section 2.3, symmetries of both types are characterized by very simple moment conditions. Table 2 allows for a precise description of the autocorrelation structure of $r_t$, $|r_t|$, $(r_t^+, r_t^-)$ for lags 1–5 and 10, 20, 40. Using the usual approximation of $1/\sqrt{T}$ as the standard error of the covariance estimates (which may be an understatement due to the nonindependence of the returns), we see that the returns appear noncorrelated. It is now well established that the stock market returns contain little serial correlation but are not independent, as shown by the significant autocorrelations generally
obtained for absolute or squared returns. In effect, the autocorrelations of the absolute values appear highly significant, which implies rejection of independence. The autocorrelation structure of \((r^-_t, r^+_t)\) is much more informative since it shows that both asymmetries are contained in the data. Although it is not a formal test, comparing lines 3 and 5 (or 4 and 6) in Table 2 reveals that the leverage effect is present, while contemporaneous asymmetry results from the fact that lines 5 and 6 are not similar.

6.1. Estimation of asymmetric GARCH models

To catch these empirical features, several GARCH-type models are fitted to the data. In accordance with many empirical studies and from a preliminary analysis, an AR(1) model is selected for the mean equation. To model the conditional variance, we limit ourselves to GARCH \((1, 1)\)-type models, i.e., low-order specifications. This choice is obviously motivated by parsimonious reasons and also by the fact that it seems sufficient to highlight the different features taken into account by the models. The BHHH algorithm is used to maximize the likelihood functions. In the first three cases, we assume that \(e_t = \sigma_t Z_t\) with \((Z_t)\) i.i.d. \(\mathcal{N}(0, 1)\).

We first estimate a symmetric volatility structure in which, following the argument of Davidian and Carroll (1987) developed in the i.i.d. case, volatility is modeled via absolute values. The estimated model (Model I in the sequel) is as follows, with estimated asymptotic standard errors given in parentheses under estimates:

\[
\begin{align*}
r_t &= 0.00046 + 0.0218 r_{t-1} + \varepsilon_t, \\
&\quad (0.00001) \quad (0.0001)
\end{align*}
\]

\[
\begin{align*}
\sigma_t &= 0.00127 + 0.143 |\varepsilon_{t-1}| + 0.777 \sigma_{t-1}. \\
&\quad (0.00010) \quad (0.000) \quad (0.000)
\end{align*}
\]

We then introduce dynamic asymmetry in the volatility structure. The following results (Model II) are obtained:

\[
\begin{align*}
r_t &= 0.00051 + 0.0255 r_{t-1} + \varepsilon_t, \\
&\quad (0.00010) \quad (0.0003)
\end{align*}
\]

\[
\begin{align*}
\sigma_t &= 0.00114 + 0.091 \varepsilon^+_{t-1} - 0.216 \varepsilon^-_{t-1} + 0.782 \sigma_{t-1}. \\
&\quad (0.00010) \quad (0.002) \quad (0.003) \quad (0.003)
\end{align*}
\]

In both models, the estimates are obtained without constraining the parameters. The results are in accordance with those provided by Rabemananjara and Zakoïan (1993) in a shorter period. The main feature is the evidence of dynamic asymmetry in the data (although not reported in the paper, Wald tests
Table 2
Sample autocorrelation functions of \( r_t, D_r_t, D_r_t \) and \( (r_t, r_t~), (D_r_t, D_r_t~) \)

<table>
<thead>
<tr>
<th>Auto correlation at lag ( k )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho(r_t, r_t~) )</td>
<td>0.030</td>
<td>0.005</td>
<td>( -0.032 )</td>
<td>0.028</td>
<td>0.005</td>
<td>0.016</td>
<td>0.003</td>
<td>0.019</td>
</tr>
<tr>
<td>( \rho(D_r_t, D_r_t~) )</td>
<td>0.000</td>
<td>0.108</td>
<td>( -0.118 )</td>
<td>0.099</td>
<td>0.009</td>
<td>0.118</td>
<td>0.055</td>
<td>0.032</td>
</tr>
<tr>
<td>( \rho(r_t, r_t~) )</td>
<td>0.028</td>
<td>0.005</td>
<td>( -0.034 )</td>
<td>0.020</td>
<td>0.003</td>
<td>0.020</td>
<td>0.003</td>
<td>0.017</td>
</tr>
<tr>
<td>( \rho(D_r_t, D_r_t~) )</td>
<td>0.013</td>
<td>0.006</td>
<td>( -0.013 )</td>
<td>0.025</td>
<td>0.002</td>
<td>0.017</td>
<td>0.002</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Indicate parameters which are statistically significant at the 5% level.

have been performed using the estimates of the covariance matrix of the parameters. As noted by many authors (see, e.g., Nelson, 1991; Ding et al., 1995), volatility tends to rise in response to excess returns lower than expected and to fall when the converse is true.

Contemporaneous asymmetry is tested by allowing two different volatilities for each component of the noise, keeping a symmetric structure for the volatilities. The estimation results (Model III) are as follows:

\[
\begin{align*}
    r_t &= 0.00035 + 0.0252 r_{t-1} + \varepsilon_t, \\
    \sigma_{t,+} &= 0.0007 + 0.115 |\varepsilon_{t-1}| + 0.426 (\sigma_{t-1,+} + \sigma_{t-1,-}), \\
    \sigma_{t,-} &= 0.0012 + 0.118 |\varepsilon_{t-1}| + 0.399 (\sigma_{t-1,+} + \sigma_{t-1,-}).
\end{align*}
\]

All the coefficients in the volatilities equations are significantly positive. The Wald tests fail to detect differences between the coefficients of $|\varepsilon_{t-1}|$ and the intercepts. Conversely, the estimated volatility coefficients are significantly different from one another.

Finally, the general asymmetric structure is estimated. Here we assume that (4) holds with $(Z_t)$ i.i.d. $\mathcal{N}(0,1)$, and the following model (IV) is obtained:

\[
\begin{align*}
    r_t &= 0.0006 + 0.0261 r_{t-1} + \varepsilon_t, \\
    \sigma_{t,+} &= 0.00094 + 0.081 \varepsilon_{t-1}^+ - 0.183 \varepsilon_{t-1}^- + 0.399 \sigma_{t-1,+} + 0.388 \sigma_{t-1,-}, \\
    \sigma_{t,-} &= 0.00354 + 0.037 \varepsilon_{t-1}^+ - 0.217 \varepsilon_{t-1}^- + 0.370 \sigma_{t-1,+} + 0.240 \sigma_{t-1,-}.
\end{align*}
\]

To check the stationarity condition, the expectations involved in Theorem 1 are obtained via simulations. The computation of the eigenvalues of the matrices involved in Theorem 1 shows that the estimated model is indeed second-order stationary. All parameters are significant at the 5% level. The estimated coefficients indicate that both types of asymmetries are present in the data. Interestingly, by comparison with the preceding model, allowing dynamic asymmetry provides a more pronounced contemporaneous asymmetry.

To assess the statistical significance of the general model, it is interesting to compare the likelihoods of the four estimated models. The results are presented in Table 3. From these results and the computation of likelihood ratio statistics, the symmetric model is rejected at any reasonable significance level. Allowing
Table 3
Estimated quasi-loglikelihoods, excess kurtosis and MSE*

<table>
<thead>
<tr>
<th>MODEL</th>
<th>loglikelihood</th>
<th>(\kappa)</th>
<th>MSE (\times 10^{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>I: SYMMETRIC</td>
<td>7385.2</td>
<td>2.49</td>
<td>0.245</td>
</tr>
<tr>
<td>II: With DYNAMIC ASYMMETRY</td>
<td>7390.6</td>
<td>2.72</td>
<td>0.280</td>
</tr>
<tr>
<td>III: With CONTEMP. ASYMMETRY</td>
<td>7396.1</td>
<td>2.57</td>
<td>0.239</td>
</tr>
<tr>
<td>IV: With ASYMMETRIES of BOTH TYPES</td>
<td>7415.5</td>
<td>2.44</td>
<td>0.187</td>
</tr>
</tbody>
</table>

*Models I–IV refer to the estimated models of Section 6.1. \(\kappa\) denotes the empirical excess kurtosis of the standardized residuals \(\hat{e}_t/\hat{\sigma}_t\). For models III and IV, \(\hat{\sigma}_t\) is calculated from (2).

Fig. 3. Estimated paths of \(\sigma_{t,+}\) (-----) and \(\sigma_{t,-}\) (- - - -) for the period Nov. 1997 to Nov. 1998.

for both asymmetries simultaneously produces a LR test statistic against model II (resp. III), of 24.9 (resp. 19.4) which has an asymptotic \(p\)-value of less than 0.001. It is also seen that the general model exhibits less excess kurtosis in the normalized residuals than the other parameterizations. However, we are not able to perform a rigorous test in this framework, and the difference might not be significant. Finally, to further compare the different specifications, we computed for each estimated model, 100 out-of-sample one-period-ahead predictions. The results are displayed in Table 3. Again, although it is not a formal test, it is seen that our general model outperforms the particular cases. These results provide strong evidence that including both types of asymmetries is statistically important as a description of the time-series properties of financial data.

Fig. 3 presents the paths of the estimated volatilities over the last year of the sample. Fig. 4 gives the path of the variable \(100(1 - \sigma_{t,+}/\sigma_{t,-})\) for the same sub-sample. It can be seen that the magnitude of the difference \((\sigma_{t,-} - \sigma_{t,+})\) often represents a significant percentage (between 5 and 15\%) of \(\sigma_{t,-}\). The residuals path is also given (see Fig. 5). It is important to note that the model has
not been constrained to produce white noise residuals (see Section 5). However, classical diagnostic tests on the estimated residuals show that the hypothesis of white noise is fully supported.

Another view of the properties of the estimated model may be obtained through the analysis of the unconditional and conditional kurtoses. Firstly, we investigated through a Monte-Carlo experiment the term structure of unconditional kurtoses, as the horizon over which returns are computed increases. The experiment involves 120 000 simulations of Model IV with i.i.d. standard Gaussian innovations $Z_t$. The kurtoses of the cumulative returns $\log(S_t/S_{t-k})$ were estimated for values of $k$ ranging from 1 to 120. The results displayed in Table 4 show that excess kurtosis is a hump-shaped function of the horizon, increasing to a maximum and then decreasing to zero for large values of $k$. This is in accordance with the results obtained by Das and Sundaram (1998) for continuous-time stochastic volatility models in the case where the Brownian motions of the mean and volatility equations are highly correlated. The sample estimates
Table 4
Term structure of the unconditional kurtosis in Model IV*

<table>
<thead>
<tr>
<th>Horizon</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kurtosis (sim.)</td>
<td>3.53</td>
<td>3.94</td>
<td>4.08</td>
<td>4.01</td>
<td>3.84</td>
<td>3.73</td>
<td>3.54</td>
<td>3.41</td>
<td>3.14</td>
<td>3.13</td>
</tr>
<tr>
<td>Kurtosis (obs.)</td>
<td>5.91</td>
<td>5.88</td>
<td>5.89</td>
<td>5.69</td>
<td>5.49</td>
<td>5.50</td>
<td>5.22</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*The simulated kurtoses are computed using a sample of 120,000 simulations of model IV. The kurtosis at horizon \( k \) is obtained for the cumulative returns \( \log(S_t/S_{t-k}) \).

Fig. 6. Estimated conditional excess kurtosis of \( \varepsilon_t \) for the period \( \frac{1197}{1198} \) to \( \frac{1200}{1201} \).

of kurtosis are also reported in Table 4 (only for values of \( k \) ranging from 1 to 30 because the kurtosis is likely to be poorly estimated for higher lags). It is clear that the simulations imply a much lower excess kurtosis over all horizons then the data set. The model has been simulated under the assumption of a conditionally Gaussian distribution, which partly explains the differences between the two-term structures. Nevertheless, at least for short horizons, the two-term structures have different shapes: the peak of kurtosis is reached for horizons of around 5 days in the model, whereas the sample kurtoses seem to be decreasing when the horizon increases. It should be noted, however, that the estimates computed over different subsets of the whole sample are highly instable, even for small horizons.

Secondly, the properties of the conditional kurtosis were investigated. Fig. 6 is a plot of the empirical conditional excess kurtosis from the general model against time. The volatilities \( \sigma_{t,+} \) and \( \sigma_{t,-} \) have been replaced by their estimated values in (3), and the moments of the (theoretically i.i.d.) normalized residuals \( Z_t = \varepsilon_t^+/\sigma_{t,+} + \varepsilon_t^-/\sigma_{t,-} \), have been replaced by their empirical counterparts on the whole sample. The most surprising feature is the great variability of this conditional excess kurtosis. By way of comparison, a nonparametric estimator
Fig. 7. Parametric (---) and nonparametric (-----) estimations of the conditional excess kurtosis of $e_t$ (100 observations starting from 25/03/89).

of the conditional excess kurtosis has been computed using the kernel method (see, e.g., Pagan and Schwert (1990) for kernel estimation of conditional variances). The conditional moments of $e_t^4$ and $e_t^2$ were calculated with a Gaussian kernel and different choices of the window width. The results appear to be very sensitive to the choice of the window width. The results reported in Fig. 7 are obtained for a bandwidth proportional to the empirical standard deviation of the data, and they show that the two estimators have close behaviours. Other computations on different subperiods lead to similar conclusions. Therefore, it seems that the parametric model IV used to estimate the first two conditional moments of $e_t^2$ is an adequate representation of the data. To investigate the behaviour of the conditional kurtosis as a function of the horizon over which returns are computed, we performed another Monte-Carlo experiment. For horizon 1, and given values of $\sigma_{t,+}$ and $\sigma_{t,-}$, the conditional kurtosis can be explicitly computed using (3). For horizon 2, the numerator is given by

$$E[\eta_{t+1}^4 | e_{t-1}] = \sum_{k=0}^{4} \binom{4}{k} E[\sigma_{t+1,+}^k \sigma_{t+1,-}^{4-k} | e_{t-1}] c(k, 4-k).$$

The conditional expectation seems difficult to obtain in closed form, but it can be evaluated by simulations. This procedure was carried out for all conditional expectations involved in the conditional kurtoses of various horizons. The results presented in Table 5 reveal that the conditional kurtoses are quite sensible to the initial values of the volatilities. When these values are equal, the conditional kurtosis at horizon 1 is that of the standard Gaussian distribution. But for higher horizons, the conditional distributions are leptokurtic, reflecting the contemporaneous asymmetry of the model. The dynamic asymmetry shows in the fact that when the initial values are inverted, the results can be very different.

6.2. Empirical evidence of asymmetries in stock returns

The empirical result that both types of asymmetries are present in the data is a typical finding in a wide range of applications. To justify this claim, we have repeated the above analysis for 42 stock returns of the French stock market. To
Table 5
Conditional kurtosis in model IV*

<table>
<thead>
<tr>
<th>( \sigma_{t,+} )</th>
<th>( \sigma_{t,-} )</th>
<th>1 day</th>
<th>5 days</th>
<th>15 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>3.00</td>
<td>3.49</td>
<td>3.46</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>3.44</td>
<td>3.31</td>
<td>3.31</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>3.42</td>
<td>3.72</td>
<td>3.68</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>3.43</td>
<td>3.28</td>
<td>3.28</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5</td>
<td>4.42</td>
<td>4.13</td>
<td>4.13</td>
</tr>
</tbody>
</table>

*This table presents the estimated conditional kurtoses at various horizons in Model IV for different initial values of \( \sigma_{t,+} \) and \( \sigma_{t,-} \). The computations involve 10,000 simulations.

Table 6
Number of rejections of the null hypothesis in LR tests at the 5% level, for 42 stocks of the CAC*

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>Alternative hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model II</td>
</tr>
<tr>
<td>Model I</td>
<td>20</td>
</tr>
<tr>
<td>Model II</td>
<td>—</td>
</tr>
<tr>
<td>Model III</td>
<td>—</td>
</tr>
</tbody>
</table>

*Numbers I–IV refer to the models in Table 3. Missing values correspond to models which are not nested.

save space the results are not reported here. They can be found in a previous draft of the paper (see El Babsiri and Zakoian, 1997). The most outstanding feature is the evidence of contemporaneous asymmetry in almost all assets, even more pronounced than the usual leverage effect. For a global comparison of the four models (Symmetric, with dynamic asymmetry, with contemporaneous asymmetric, with both asymmetries) the normal quasi-likelihoods are used to compute LR statistics. This selection criterion chooses the general asymmetric model for a very large majority of assets (see Table 6). For only five stocks, including the leverage effect in addition to contemporaneous asymmetry appears unnecessary. The table also confirms that contemporaneous asymmetry is more important in the data than the classical leverage effect. To evaluate how dynamic asymmetry affects the two volatility processes as a short term effect, it is worthwhile comparing the coefficients of \( e_{t-1}^+ \) and \( e_{t-1}^- \). Table 7 summarizes the results. For almost one out of four assets, the volatilities exhibit symmetric reactions to past positive and negative shocks. One striking feature is that the volatility of negative changes overreacts to past negative changes (compared to positive ones) for about half of the sample (it is symmetric in the other cases). On the other hand, the volatility of positive changes provides a more contrasted
Table 7
News impact sets for the 42 stocks of the CAC index

<table>
<thead>
<tr>
<th>No. of items</th>
<th>( a_{i,+}^\dagger = a_{i,-} )</th>
<th>( a_{i,+}^\dagger &gt; a_{i,-} )</th>
<th>( a_{i,+}^\dagger &lt; a_{i,-} )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_{i,-} = a_{i,-} )</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>( a_{i,-} &gt; a_{i,-} )</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>( a_{i,-} &lt; a_{i,-} )</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>15</td>
<td>9</td>
<td>42</td>
</tr>
</tbody>
</table>

Each cell describes the number of stocks corresponding to a particular case of asymmetry (or symmetry) in the volatility equations (e.g., there are seven stocks such that: (i) positive values of \( a_{i,-} \) have bigger impacts on \( a_{i,+} \) than negative values of the same magnitude; (ii) the converse is true for \( a_{i,-} \)).

6.3. Comparison with alternative approaches

It is interesting to compare these results with the alternative models proposed by Hansen (1994). As mentioned in Section 2, these specifications have been designed to model non-Gaussian time-varying distributions (instead of asymmetries). The first step is to consider a family of density functions, noted \( g(\cdot | \eta, \lambda) \), generalizing the standard normal. The explicit form is given in Appendix C. The interpretation of \( \eta \) is similar to the degrees of freedom in the Student’s \( t \) density, whereas \( \lambda > 0 \) (resp. \( < 0 \)) indicates that the mode of the density is to the left (resp. right) of zero and the variable is skewed to the right (resp. left). Then we estimate a symmetric volatility structure and (omitting the conditional mean equation, which is very similar to those already obtained) we get

\[
\sigma_t = 0.00105 + 0.120 |a_{t-1}| + 0.751 \sigma_{t-1}.
\]

\[
(0.04190) \quad (0.360) \quad (0.137)
\]

Moreover, a \( g(\cdot | 3.63, 0.72) \) density function is obtained for the unconditional distribution. As a comparison with the asymmetric models, the estimated loglikelihood is equal to 7408.2. So allowing for a more general density than the normal significantly increases the log-likelihood. We next allowed the degrees-of-freedom parameter to be time-varying. We used the same specifications as Hansen (1994) to constrain the time-varying conditional degrees-of-freedom parameters. The estimated conditional variance equation is almost identical...
to (14), so we do not report it. The estimations obtained for the time-varying degrees of freedom are presented in Appendix C. The results of the fitting suggest a higher loglikelihood (7524.5) as for our asymmetric model (7415.5). However, it is difficult to be rigorous in this comparison as the models are not nested, which means that basing the usual likelihood ratio test on asymptotic $\chi^2$ distributions is not theoretically founded. Moreover, several estimated coefficients are non significant at any reasonable level. Finally, the MSEs are, respectively, equal to 0.000192 and 0.000172 for the Hansen’s models. These results lead to the same conclusions as the log-likelihood comparison.

We also found interesting to compare our formulation with a discretized version of a continuous-time stochastic volatility model with correlated Brownian motions. Maximum likelihood estimation of the model proposed by Heston and Nandi (1999) provides the following results:

$$
\begin{align*}
  r_t &= 0.0003 + 0.043 \sigma_t^2 + \sigma_t Z_t, \\
  \sigma_t^2 &= 0.002 + 0.769 \sigma_{t-1}^2 + 0.022 [Z_{t-1} - 1.24\sigma_{t-1}]^2,
\end{align*}
$$

where $(Z_t)$ is an i.i.d. $\mathcal{N}(0, 1)$ process. The existence of a significant coefficient for $\sigma_{t-1}$ in the volatility equation confirms the presence of dynamic asymmetry in the data. The estimated log-likelihood is equal to 7395.8, which is similar to the values obtained for Model III. However, because the models are not nested it is difficult to draw definitive conclusions. It should be noted, however, that the Heston and Nandi model is unable to capture the contemporaneous asymmetry property since the term $\sigma_t Z_t$ in the return equation is a martingale difference.

Another way to compare stochastic variance models is to consider their ability to predict option prices. A first step in this direction has been achieved by Härdle and Hafner (2000), who compared several GARCH-type models in terms of option pricing. They have used out-of-the-money calls on the German stock index DAX. A remarkable finding is that the threshold GARCH model works significantly better than the other models. Concerning the option pricing with our general formulation, and the comparison with continuous time models allowing for conditional skewness and kurtosis (see, e.g., Das and Sundaram, 1998), it is clear that the work made by Duan (1995), Kallsen and Taqqu (1998) on standard GARCH models needs to be adapted. To compare the empirical performances of the different option pricing models, the approach of Bakchi et al. (1998) provides useful tools. We leave this issue for future researches.
7. Conclusion

In this paper, we introduced and illustrated a new concept of asymmetry in the dynamics of stock returns. Not only do the signs of past shocks affect the current volatility, but also, the impact depends on whether the current excess return is lower or higher than expected. This was achieved by relaxing the classical martingale difference assumption on GARCH innovation processes. With this model, we are able to capture some effects such as contemporaneous asymmetry, heteroskedawness and heterokurtosis. An important feature of the proposed model is that, enlarging the class of GARCH-type models, it keeps their tractability: due to its linearity properties, it allows for Markov representations useful to analyze the probability structure; moreover, the statistical procedures currently used in the GARCH framework (two-stage least squares, QMLE) remain consistent under some identifiability assumptions. In addition, the empirical study suggested that the underlying assumptions of the model (asymmetry of both types, time-varying conditional skewness and kurtosis) can be strongly supported by financial data. Comparison with alternative specifications shows that our model dominates formulations that do not allow for heteroskedawness, heterokurtosis and asymmetries. Another source of information concerning the conditional variance is of course options data. The model of this paper provides a framework for introducing the different implicit volatilities of calls and puts observed by practitioners. A direction of future research could be to consider these implicit volatilities in order to improve option pricing and hedging.

Acknowledgements

J.-M. Zakoian gratefully acknowledges financial support from the HCM program. We would like to thank Adel Grar for providing us with the estimation program used in the empirical section, and Ali Alami for computational assistance. We are also grateful for the helpful comments and suggestions from an associate editor and two referees.

Appendix A. Proofs of the results of Section 4

Proof of Theorem 1. We use a similar approach as in Liu and Brockwell (1988). Let us first define the following $\mathbb{R}^{2(p+q-1)}$-valued stochastic processes:

$$X_n(t) = \begin{cases} 0 & \text{if } n < 0, \\ b + C(Z_t)X_{n-1}(t-1) & \text{if } n \geq 0 \end{cases}$$
and for all \( n \in \mathbb{Z} \),

\[
W_n(t) = X_n(t) - X_{n-1}(t), \quad t \in \mathbb{Z}.
\]

It is easily seen that for all \( n > 0 \), \( X_n(t) \) and \( W_n(t) \) are measurable functions of \( Z_t, Z_{t-1}, \ldots, Z_{t-n+1} \). As a consequence, the two processes are strictly stationary (for fixed \( n \)) and the couple \( \{X_n(t), W_n(t), t \in \mathbb{Z}\} \) is also strictly stationary. Moreover, from the definitions of \( X_n(t) \) and \( W_n(t) \) we have

\[
W_n(t) = \begin{cases} 
0 & \text{if } n < 0, \\
b & \text{if } n = 0, \\
C(Z_t)W_{n-1}(t-1) & \text{if } n > 0.
\end{cases}
\]

It is then easy to show, by iterating the preceding relation and from the independence of matrices \( C(Z_t) \), that the expectation of \( W_n(t) \) is finite. Therefore we can set for all \( n \in \mathbb{Z} \):

\[
\Gamma_n = E[W_n(t)W_n'(t)] - E[W_n(t)]E[W_n'(t)]
\]

\[= E[C(Z_t)W_{n-1}(t-1)W_{n-1}'(t-1)C'(Z_t)]
\]

\[= E[C(Z_t)W_{n-1}(t-1)W_{n-1}'(t-1)C'(Z_t)].
\]

Using the independence between \( Z_t \) and \( W_{n-1}(t-1) \) we then have

\[
\text{Vec} \Gamma_n = C^* \text{Vec} \{E[W_{n-1}(t-1)W_{n-1}'(t-1)]\}
\]

\[+ (\tilde{C} \otimes \tilde{C}) \text{Vec} \{E[W_{n-1}(t-1)]E[W_{n-1}'(t-1)]\}
\]

\[= C^* \text{Vec} \Gamma_{n-1} + [C^* + \tilde{C} \otimes \tilde{C}]
\]

\[
\text{Vec} \{E[W_{n-1}(t-1)]E[W_{n-1}'(t-1)]\}.
\]

Moreover,

\[
\text{Vec} \{E[W_n(t)]E[W_n'(t)]\} = (\tilde{C} \otimes \tilde{C}) \text{Vec} \{E[W_{n-1}(t-1)]E[W_{n-1}'(t-1)]\}.
\]

Finally, if we set

\[
\Lambda_n = [(\text{Vec} \Gamma_n)'(\text{Vec} \{E[W_n(t)]E[W_n'(t)]\})']
\]

and

\[
\Sigma = \begin{bmatrix} C^* & C^* + \tilde{C} \otimes \tilde{C} \\ 0 & \tilde{C} \otimes \tilde{C} \end{bmatrix},
\]

we have \( \Lambda_n = \Sigma \Lambda_{n-1} \). Since the eigenvalues of \( \tilde{C} \otimes \tilde{C} \) are the products of those of \( \tilde{C} \), we can conclude that \( \rho(\Sigma) < 1 \).
Let \(|A_n|\) denote the sum of the absolute values of all the components of \(A_n\). From Liu and Brockwell (1988, Proposition 2.1), we have the inequality \(|A_n| \leq \text{const. } \rho(\Sigma)^n\) at least for large \(n\), which ensures that the \(L^2\) norm of \(\{X_n(t) - X_{n-1}(t)\}\) decreases at a geometric rate. By the Cauchy criterion for the convergence in \(L^2\), the sequence \(\{X_n(t)\}\) converges in mean square for each \(t\), to some limit \(X_t\), say. It is easy to see that the limit process is in fact strictly stationary and can be expressed in terms of present and past values of the \(\{Z_t\}\) process (i.e., \(\{Z_t\}\) is causal). Therefore, it is also ergodic.

To prove uniqueness, let \(\tilde{X}_t\) denote another second-order stationary causal solution of (8) and let \(\delta_t = X_t - \tilde{X}_t\). We have \(\delta_{t+1} = C(Z_{t+1})\delta_t\). Let

\[
\hat{R} = \begin{bmatrix}
\text{Vec}(E(\delta_t\delta_t')) - (E\delta_t)(E\delta_t') \\
\text{Vec}(E(\delta_t')(E\delta_t'))
\end{bmatrix}
\]

From the above discussion we have \(\hat{R} = \Sigma\hat{R} = \lim_{n \to \infty} \Sigma^n \hat{R} = 0\), which proves uniqueness of \(\{X_t\}\) in \(L^2\).

Finally, \(\{e_t\}\) is ergodic and (weakly and strictly) stationary as the sum of two components of \(\{X_t\}\). It is also unique in \(L^2\) since there is a one-to-one correspondence between \(\{X_t\}\) and \(\{e_t\}\) for a given \(\{Z_t\}\) process. \(\square\)

**Proof of Corollary 1.** From (8) we have

\[
\text{Vec}(X_{t+1}X'_{t+1}) = [C(Z_{t+1}) \otimes C(Z_{t+1})] \text{Vec}(X_tX'_t)
\]

\[+ [b \otimes C(Z_{t+1}) + C(Z_{t+1}) \otimes b] X_t + \text{Vec}(bb').\]

Therefore, \(E[W_{t+1}] = A\hat{E}[W_t] + \hat{b}\). The result holds since \(W_t\) is second-order stationary and \(I - \hat{A}\) is invertible. \(\square\)

**Appendix B. Proofs of the results of Section 5**

**Proof of Theorem 2.** To establish the consistency of the QMLE, it will be useful to introduce the (unobserved) objective function

\[
L^*_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ell^*_t(\theta),
\]

obtained by replacing \(\{e_t(\theta), s_{t,+}(\theta), s_{t,-}(\theta)\}\) by \(\{\hat{e}_t(\theta), \sigma_{t,+}(\theta), \sigma_{t,-}(\theta)\}\). For all \(\theta \in \Theta_d\), let \(v^*_t(\theta) = \sigma^2_{t,+}(\theta)1_{e_t(\theta) > 0} + \sigma^2_{t,-}(\theta)1_{e_t(\theta) < 0}\) and let \(v^*_t = v^*_t(\theta_0)\). We then have

\[
\ell^*_t(\theta) = -\frac{1}{2} \log v^*_t(\theta) - \frac{\hat{e}^2_t(\theta)}{2v^*_t(\theta)}.
\]
The advantage of using $L^T(\theta)$ instead of $L_T(\theta)$ is that it is based on stationary ergodic sequences. The first lemma justifies the approximation of $L_T(\theta)$ by $L^T(\theta)$. □

**Lemma B.1.** We have almost surely

$$\lim_{T \to \infty} \sup_{\theta \in \Theta} |L_T(\theta) - L^T(\theta)| = 0.$$ 

**Proof.** (i) We first show that $\lim_{t \to \infty} \sup_{\theta \in \Theta} |\epsilon_t(\theta) - \bar{\epsilon}_t(\theta)| = 0$.

For $Q = 0$ the result is obvious. Otherwise, let us introduce the square matrix

$$B = \begin{pmatrix} -\psi_1 & -\psi_2 & \ldots & -\psi_{Q-1} & -\psi_Q \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}$$

and the $Q \times 1$ vectors $e_t(\theta) = (e_t(\theta), \ldots, e_{t-Q+1}(\theta))'$ and $e_t(\theta) = (e_t(\theta), \ldots, e_{t-Q+1}(\theta))'$. We have $e_t(\theta) - \bar{e}_t(\theta) = B^{-t}(e_P(\theta) - e_P(\theta)), \forall t \geq P$. Since the nonzero eigenvalues of $B$ are the inverses of those of the polynomial $\Psi(z)$, the spectral radius of $B$ is less than $(1 + \delta)^{-1}$ for all $\theta \in \Theta$. Therefore, there exists a positive constant $c$, independent of $\theta$, such that $\|B^t\| \leq c(t(1 + \delta))^{-t}, \forall t > 0$, where $\| \cdot \|$ denotes any matrix norm (see Francq and Zakoian, 1998). The conclusion follows.

(ii) Then we show that

$$\lim_{t \to \infty} \sup_{\theta \in \Theta} |\sigma_{t,+}(\theta) - s_{t,+}(\theta)| = \lim_{t \to \infty} \sup_{\theta \in \Theta} |\sigma_{t,-}(\theta) - s_{t,-}(\theta)| = 0.$$ 

We have

$$\begin{pmatrix} \sigma_{t,+}(\theta) - s_{t,+}(\theta) \\ \sigma_{t,-}(\theta) - s_{t,-}(\theta) \end{pmatrix} = \left[ I - B_\theta(L) \right]^{-1} A_\theta(L) \begin{pmatrix} e^+_t(\theta) - e^+_t(\theta) \\ e^-_t(\theta) - e^-_t(\theta) \end{pmatrix},$$

(B.3)

where all missing terms are replaced by zeros. It is clear from (i) that the terms $e^+_t(\theta) - e^+_t(\theta)$ and $- e^-_t(\theta) + e^-_t(\theta)$ converge uniformly to zero as $t$ goes to infinity. Then the conclusion follows from similar argument as in (i) since the inverses of the zeros of the polynomial $I - B_\theta(z)$ are uniformly bounded by $(1 + \delta)^{-1}$. 

(iii) Finally, we have

\[
\lim_{t \to \infty} \sup_{\theta \in \Theta_s} |\ell_t(\theta) - \ell_t^s(\theta)|
\]

\[
= \lim_{t \to \infty} \sup_{\theta \in \Theta_s} \left| \frac{1}{2} \log \frac{v_t^*(\theta)}{v_t(\theta)} + \frac{e_t^2(\theta)}{2v_t^*(\theta)} - \frac{e_t^2(\theta)}{2v_t(\theta)} \right|
\]

\[
\leq \lim_{t \to \infty} \sup_{\theta \in \Theta_s} \left| \frac{1}{2} \log \frac{v_t^*(\theta)}{v_t(\theta)} \right| + \lim_{t \to \infty} \sup_{\theta \in \Theta_s} \left| \frac{e_t^2(\theta)}{2v_t^*(\theta)} - \frac{e_t^2(\theta)}{2v_t(\theta)} \right|
\]

where \(v_t(\theta)\) denotes the observable counterpart of \(v_t^*(\theta)\). Now, we note that

\[
\sup_{\theta \in \Theta_s} \left| \frac{v_t^*(\theta)}{v_t(\theta)} \right| = \sup_{\theta \in \Theta_s} \left| 1 + \frac{v_t^*(\theta) - v_t(\theta)}{v_t(\theta)} \right|
\]

\[
\leq 1 + \frac{1}{\delta} \sup_{\theta \in \Theta} |v_t^*(\theta) - v_t(\theta)|,
\]

and the same inequality holds for \(\sup_{\theta \in \Theta_s} |v_t(\theta)/v_t^*(\theta)|\) since, from the assumptions, \(v_t(\theta)\) (resp. \(v_t^*(\theta)\)) is uniformly bounded away from zero. From (i), it is clear that \(\sup_{\theta \in \Theta_s} |v_t^*(\theta) - v_t(\theta)|\) converges to zero as \(t\) goes to infinity. Moreover, we have

\[
\left| \frac{e_t^2(\theta)}{2v_t^*(\theta)} - \frac{e_t^2(\theta)}{2v_t(\theta)} \right| \leq \frac{1}{2v_t^*(\theta)} |e_t^2(\theta) - e_t^2(\theta)| + \frac{e_t^2(\theta)}{2v_t(\theta)v_t^*(\theta)} |v_t^*(\theta) - v_t(\theta)|
\]

\[
\leq \frac{1}{2\delta} |e_t^2(\theta) - e_t^2(\theta)| + \frac{1}{2\delta^2} |v_t^*(\theta) - v_t(\theta)| e_t^2(\theta).
\]

Both terms on the right-hand side of the previous equation go to zero from (i), (ii) and the fact that \(e_t(\cdot)\) and \(e_t(\cdot)\) are continuous functions, hence bounded, on \(\Theta_s\). This completes the proof. \(\square\)

We now prove the following identifiability result.

**Lemma B.2.** Let \(\theta \in \Theta_s\) and suppose that \(\forall t \in \mathbb{Z}\),

\[
e_t(\theta) = e_t \quad (a.s.), \quad \sigma_t(\theta) = \sigma_t(\theta) \quad (a.s.), \quad \sigma_t(\theta) = \sigma_t(\theta) \quad (a.s.). \quad (B.4)
\]

Then \(\theta = \theta_0\).

**Proof.** Let us partition the parameter vector into \(\theta' = (\theta^{(1)'}\, , \theta^{(2)'})\), where \(\theta^{(1)}\) (resp. \(\theta^{(2)}\)) denotes the coefficients of the conditional mean (resp. volatilities) equation.
(i) First, we have $\forall \theta \in \Theta_{\Delta}, \forall t \in \mathbb{Z}, e_i(\theta) = \varepsilon_i$ implies $\theta^{(1)} = \theta_0^{(1)}$. Otherwise, one could find a linear combination equal to zero of $Y_{t-1}, Y_{t-2}, \ldots$. This is impossible from the assumption that $\{Y_t\}$ is a regular process.

(ii) From the invertibility assumptions on the lag polynomials in (11), for all $\theta \in \Theta_{\Delta}$, $\sigma_{i,+}(\theta)$ and $\sigma_{i,-}(\theta)$ can be written as linear combinations of the infinite past of $(\varepsilon_i^+(\theta), \varepsilon_i^-(\theta))$:

$$\sigma_{i,+}(\theta) = c_{i,+}(\theta) + \sum_{i=1}^{\infty} c_{i,+}(\theta)e_{t-i}^+(\theta) + d_{i,+}(\theta)e_{t-i}^-(\theta),$$

$$\sigma_{i,-}(\theta) = c_{i,-}(\theta) + \sum_{i=1}^{\infty} c_{i,-}(\theta)e_{t-i}^+(\theta) + d_{i,-}(\theta)e_{t-i}^-(\theta),$$

where the sequences $(c_{i,+}(\theta)), (c_{i,-}(\theta)), (d_{i,+}(\theta)), (d_{i,-}(\theta))$ are absolutely summable.

Now, let $\theta \in \Theta_{\Delta}$ with $\{e_i(\theta), \sigma_{i,+}(\theta), \sigma_{i,-}(\theta)\} = \{\varepsilon_i, \sigma_{i,+}, \sigma_{i,-}\}, \forall t \in \mathbb{Z}$. We have

$$c_{0,+}(\theta) + \sum_{i=1}^{\infty} c_{i,+}(\theta)e_{t-i}^+ + d_{i,+}(\theta)e_{t-i}^- = c_{0,+}(\theta_0) + \sum_{i=1}^{\infty} c_{i,+}(\theta_0)e_{t-i}^+ + d_{i,+}(\theta_0)e_{t-i}^-.$$

Hence, $(c_{1,+}(\theta) - c_{1,+}(\theta_0))\sigma_{i,-1,+} + Z_{t-1,+} + (d_{1,+}(\theta) - d_{1,+}(\theta_0))\sigma_{i,-1,-} - Z_{t-1,-}$ can be written as a function of $e_{t-2}, e_{t-3}, \ldots$. Since $Z_{t-1}$ is independent of the past of $e_{t-2}$ we have: $c_{i,+}(\theta) = c_{i,+}(\theta_0), d_{i,+}(\theta) = d_{i,+}(\theta_0)$. Iterating the procedure, we show that: $\forall i \geq 0, c_{i,+}(\theta) = c_{i,+}(\theta_0), d_{i,+}(\theta) = d_{i,+}(\theta_0)$. Similarly, we have $(c_{i,-}(\theta)) = (c_{i,-}(\theta_0)), (d_{i,-}(\theta)) = (d_{i,-}(\theta_0))$. Therefore, we have proved that

$$[I - B_{\theta}(z)]^{-1}A_{\theta}(z) = [I - B_{\theta_0}(z)]^{-1}A_{\theta_0}(z), \forall z, ||z|| \leq 1.$$

The conclusion follows from the fact that the orders of the polynomials $B_{\theta}$ and $B_{\theta_0}$ (resp. $A_{\theta}$ and $A_{\theta_0}$) are the same, and that $B_{\theta}$ and $A_{\theta}$ (resp. $B_{\theta_0}$ and $A_{\theta_0}$) have no common root.

Asymptotic identifiability is a consequence of the next lemma.

**Lemma B.3.** For all $\theta \in \Theta_{\Delta}$, let $L^*(\theta) = E_{\theta_0}(\ell^*_t(\theta))$. Under the assumptions of Theorem 2, $L^*(\theta)$ exists and attains a unique maximum at $\theta_0$.

**Proof.** (i) We first prove that $\forall \theta \in \Theta_{\Delta}, -\infty < L^*(\theta) < +\infty$. 

The assumption on the MA polynomial $\Psi_\theta$ implies that there exists a sequence of absolutely summable constants $(c_i(\theta))$ such that $\forall t \in \mathbb{Z}$, $\varepsilon_t(\theta) = Y_t + \sum_{i=1}^{t} c_i(\theta) Y_{t-i}$. Therefore, $\{\varepsilon_t(\theta)\}$ belongs to $L^2$, $\forall \theta \in \Theta_\delta$. In a similar way, the variables $\sigma_{i,+}(\theta)$ and $\sigma_{i,-}(\theta)$ belong to $L^2$ (from invertibility of the lag polynomials). Now, we have

$$-2L^*(\theta) = E_{\theta_0} \log[v^*_{i}(\theta)] + E_{\theta_0} \left[ \frac{\varepsilon_i^2(\theta)}{v^*_{i}(\theta)} \right].$$

From the definition of $\Theta_\delta$, it is clear that $v^*_{i}(\theta) \geq \delta^2 > 0$. Therefore, $E_{\theta_0} \log[v^*_{i}(\theta)] \geq 2 \log \delta > -\infty$. Moreover, $E_{\theta_0} \log[v^*_{i}(\theta)] < E_{\theta_0} \log[v^*_{i}(\theta)] < E_{\theta_0} [\sigma_{i,+}^2(\theta) + \sigma_{i,-}^2(\theta)] < +\infty$. Finally, $E_{\theta_0} [\varepsilon_i^2(\theta)/v^*_{i}(\theta)] \leq (1/\delta^2)$ $E_{\theta_0} \varepsilon_i^2(\theta) < +\infty$.

(ii) Now let

$$\Delta(\theta) = 2(\log[v^*_{i}(\theta)] - \log[v^*_{i}(\theta_0)])$$

$$= E_{\theta_0} \log \left( \frac{v^*_{i}(\theta)}{v^*_{i}(\theta_0)} \right) + \left( \frac{\varepsilon_i^2(\theta)}{v^*_{i}(\theta)} - \frac{\varepsilon_i^2(\theta_0)}{v^*_{i}(\theta_0)} \right)$$

$$= E_{\theta_0} \log \left( \frac{v^*_{i}(\theta)}{v^*_{i}(\theta_0)} \right) + Z_i^2 \left( 1 - \frac{v^*_{i}(\theta)}{v^*_{i}(\theta_0)} \right) + \frac{\varepsilon_i^2(\theta) - \varepsilon_i^2(\theta_0)}{v^*_{i}(\theta)}$$

$$= E_{\theta_0} \left[ \log \left( \frac{v^*_{i}(\theta)}{v^*_{i}(\theta_0)} \right) + 1 - \frac{v^*_{i}(\theta)}{v^*_{i}(\theta_0)} \right]$$

$$+ \left( Z_i^2 - 1 \right) \left( 1 - \frac{v^*_{i}(\theta)}{v^*_{i}(\theta_0)} \right) + \frac{\varepsilon_i^2(\theta) - \varepsilon_i^2(\theta_0)}{v^*_{i}(\theta)}.$$

Suppose that assumption (i) of Theorem 2 holds. Then $v^*_{i}(\theta) \in \varepsilon_{i-1}, \forall \theta \in \Theta_\delta$ and therefore $E_{\theta_0} [(Z_i^2 - 1)(1 - \frac{v^*_{i}(\theta)}{v^*_{i}(\theta_0)})] = 0$. Moreover, we have

$$E_{\theta_0} \left[ \frac{\varepsilon_i^2(\theta)}{v^*_{i}(\theta)} \right] = E_{\theta_0} \left[ \frac{2 \varepsilon_i (\varepsilon_i - \varepsilon_i(\theta))}{v^*_{i}(\theta)} - \frac{(\varepsilon_i - \varepsilon_i(\theta))^2}{v^*_{i}(\theta)} \right]$$

$$= -E_{\theta_0} \left[ \frac{(\varepsilon_i - \varepsilon_i(\theta))^2}{v^*_{i}(\theta)} \right] \leq 0$$

since $\{\varepsilon_i\}$ is a martingale difference in this case and $\varepsilon_i - \varepsilon_i(\theta) \in \varepsilon_{i-1}$.

Now, if assumption (ii) holds, we have $v^*_{i}(\theta) = \sigma_{i,+}^2(\theta)1_{Z_i > 0} + \sigma_{i,-}^2(\theta)1_{Z_i < 0}$ and since $(Z_i)$ is independent of all past variables it is easily obtained by conditioning that $\Delta(\theta) \leq 0$. 


When \( \Delta(\theta) = 0 \), we have \( v_i^e(\theta) = v_i^e \) and \( \varepsilon_i(\theta) = \varepsilon_i \). Therefore, from Lemma B.2 we have \( \theta = \theta_0 \).

**Lemma B.4.** For all \( \Theta_\delta \), \( \theta_1 \neq \theta_0 \), there exists a.s. a neighbourhood \( V(\theta_1) \) of \( \theta_1 \), \( V(\theta_1) \subset \Theta_\delta \), such that

\[
\limsup_{T \to \infty} \sup_{\theta \in V(\theta_1)} L_T^*(\theta) < L_T^*(\theta_0).
\]

**Proof.** Following the approach of Francq and Zakoïan (1998), let \( V_m(\theta_1) \) denote the sphere of radius \( 1/m \) and centered at \( \theta_1 \). Let \( S_m(t) = \sup_{\theta \in V_m(\theta_1)} \ell_T^*(\theta) \). As well as \( \ell_T^*(\theta) \), \( S_m(t) \) belongs to \( L^1 \). The ergodic theorem shows that a.s.,

\[
\sup_{\theta \in V_m(\theta_1)} L_T^*(\theta) \leq \frac{1}{T} \sum_{t=1}^{T} S_m(t) \to E_{\theta_0} S_m(t),
\]

as \( T \) goes to infinity. Since \( \ell_T^*(\theta) \) is a smooth function of \( \theta \), \( S_m(t) \) decreases to \( \ell_T^*(\theta_1) \) as \( m \) goes to infinity. Therefore, from Lebesgue’s theorem we have

\[
\lim_{m \to \infty} E_{\theta_0} S_m(t) = L_T^*(\theta_1).
\]

By Lemma B.3, we can write that

\[
\limsup_{m \to \infty} \limsup_{T \to \infty} \sup_{\theta \in V_m(\theta_1)} L_T^*(\theta) \leq L_T^*(\theta_1) < L_T^*(\theta_0).
\]

The conclusion follows. \( \square \)

**Lemma B.5.** For all \( \Theta_\delta \), \( \theta_1 \neq \theta_0 \), there exists a.s. a neighbourhood \( V(\theta_1) \) of \( \theta_1 \), \( V(\theta_1) \subset \Theta_\delta \), such that

\[
\limsup_{T \to \infty} \sup_{\theta \in V(\theta_1)} L_T(\theta) < L_T^*(\theta_0).
\]

**Proof.** The result is straightforward from Lemmas B.1 and B.4, since

\[
\sup_{\theta \in V(\theta_1)} L_T(\theta) \leq \sup_{\theta \in V(\theta_1)} L_T^*(\theta) + \sup_{\theta \in V(\theta_1)} |L_T(\theta) - L_T^*(\theta)|.
\]

Let \( V(\theta_0) \) be any neighbourhood of \( \theta_0 \). For all \( \theta_1 \in \Theta_\delta - V(\theta_0) \), there exists a.s. a neighbourhood \( V(\theta_1) \) such as in Lemma B.5. Since \( \Theta_\delta \) is compact, there exist \( \theta_1, \ldots, \theta_k \) such that \( \Theta_\delta \) is covered by \( V(\theta_0), V(\theta_1), \ldots, V(\theta_k) \). Lemma B.5 shows that, a.s.,

\[
\sup_{\theta \in \Theta_\delta} L_T(\theta) = \max_{i=0, 1, \ldots, k} \sup_{\theta \in V(\theta_i)} L_T(\theta) = \sup_{\theta \in V(\theta_i)} L_T(\theta)
\]

for \( T \) large enough. Therefore, the QML estimator almost surely belongs to \( V(\theta_0) \) for large \( T \), which completes the proof of Theorem 2.
Proof of Theorem 3. Since the proof is very similar to that of Theorem 2, we will not give it to save space. However, we proof that the expected quasi-loglikelihood has a unique maximum at the true parameter value. For ease of exposition, we assume that the variances of $Z^+$ and $Z^-$, which are known in this theorem, are both equal to 1. Let

$$
\tilde{\mathcal{L}}_{\theta}^*(\theta) = -\left(\log \sigma_i(\theta) + \frac{\eta_i^2(\theta)}{2\sigma_i^2(\theta)}\right),
$$

where

$$
\eta_i(\theta) = \varepsilon_i(\theta) - \sigma_{i,+}(\theta) + \sigma_{i,-}(\theta),
$$

$$
\sigma_i(\theta)^2 = \sigma_{i,+}(\theta) + \sigma_{i,-}(\theta) + 2\sigma_{i,+}(\theta)\sigma_{i,-}(\theta),
$$

with $\eta_i(\theta_0) = \eta_i$, $\sigma_i(\theta_0) = \sigma_i$. From (9), we have $\eta_i = \sigma_i u_i$.

Lemma B.6. For all $\theta \in \Theta_\beta$, let $\tilde{\mathcal{L}}^*(\theta) = \mathbb{E}_{\theta_0}(\tilde{\mathcal{L}}^*(\theta))$. Under the assumptions of Theorem 3, $\tilde{\mathcal{L}}^*(\theta)$ exists and attains a unique maximum at $\theta_0$.

Proof. (i) By similar arguments as in Lemma B.4 it can be shown that $\forall \theta \in \Theta_\beta, -\infty < \tilde{\mathcal{L}}^*(\theta) < +\infty$.

(ii) Let

$$
\tilde{\Lambda}(\theta) = 2(\tilde{\mathcal{L}}^*(\theta) - \tilde{\mathcal{L}}^*(\theta_0))
$$

$$
= \mathbb{E}_{\theta_0} \left[ \log \left( \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} \right) + \left( \frac{\eta_i^2(\theta)}{\sigma_i^2(\theta)} - \frac{\eta_i^2(\theta_0)}{\sigma_i^2(\theta_0)} \right) \right]
$$

$$
= \mathbb{E}_{\theta_0} \left[ \log \left( \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} \right) + 1 - \frac{\eta_i^2(\theta)}{\sigma_i^2(\theta_0)} + (u_i^2 - 1) \left( 1 - \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} \right) ight.
$$

$$
+ \left. \frac{\eta_i^2(\theta) - \eta_i^2(\theta_0)}{\sigma_i^2(\theta_0)} \right]
$$

$$
= \mathbb{E}_{\theta_0} \left[ \log \left( \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} \right) + 1 - \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} + (u_i^2 - 1) \left( 1 - \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} \right) ight]
$$

$$
+ \frac{2\eta_i(\theta) - \eta_i(\theta_0)}{\sigma_i^2(\theta_0)} - \frac{(\eta_i - \eta_i(\theta_0))^2}{\sigma_i^2(\theta_0)} \right].
$$

We have

$$
\mathbb{E}_{\theta_0} \left[ \log \left( \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} \right) + 1 - \frac{\sigma_i^2(\theta)}{\sigma_i^2(\theta_0)} \right] \leq 0.
$$

(B.5)
Moreover, \( \forall \theta \in \Theta_{\delta}, \sigma^2_t(\theta) \in \varepsilon_{t-1} \). Therefore, by taking the expectation conditional to the past:

\[
E_{\theta_0} \left[ (u_t^2 - 1) \left( 1 - \frac{\sigma_t^2}{\sigma^2_t(\theta)} \right) \right] = 0.
\]

Finally, since \( \forall \theta \in \Theta_{\delta}, \eta_t = \eta_t(\theta) \in \varepsilon_{t-1}, \sigma_t^2(\theta) \in \varepsilon_{t-1} \) and since \( \{\eta_t\} \) is a martingale difference:

\[
E_{\theta_0} \left[ \frac{2\eta_t(\eta_t - \eta_t(\theta))}{\sigma_t^2(\theta)} \right] = 0.
\]

Therefore, we have proved that \( \forall \theta \in \Theta_{\delta}, A(\theta) = 0 \).

(iii) Now suppose that \( A(\theta) = 0 \). Then the terms into brackets in (19) is a.s. equal to zero, which proves that \( \sigma_t(\theta)^2 = \sigma_t^2 \), a.s. Therefore

\[
\sigma_{t,+}(\theta) + \sigma_{t,-}(\theta) = \sigma_{t,+} + \sigma_{t,-}, \quad \text{a.s.}
\]

Since \( \forall \theta \in \Theta_{\delta}, \sigma_{t,+} > 0, \sigma_{t,-} > 0 \). Moreover, we have

\[
E_{\theta_0} \left[ \frac{(\eta_t - \eta_t(\theta))^2}{\sigma_t^2(\theta)} \right] = 0
\]

which implies that \( \eta_t = \eta_t(\theta) \), a.s. Finally, we have: \( \varepsilon_t - \varepsilon_t(\theta) = 2(\sigma_{t,+} - \sigma_{t,+}(\theta)) \).

Now, if we had \( \varepsilon_t - \varepsilon_t(\theta) \neq 0 \), \( Y_{t-1} \) could be expressed as a function of \( Y_{t-2}, Y_{t-3}, \ldots \). Therefore, the innovation of \( Y_{t-1} \) would be zero which is excluded since \( \{Y_t\} \) is a non deterministic process. Therefore, \( \varepsilon_t = \varepsilon_t(\theta) \) and then \( \sigma_{t,+} = \sigma_{t,+}(\theta) \) and \( \sigma_{t,-} = \sigma_{t,-}(\theta) \). Finally, from Lemma B.2, \( \theta = \theta_0 \). Therefore, the result is proved. \( \square \)

Appendix C. Alternative conditional distributions used in Section 6

Following Hansen (1994), consider the density function (normalized to have zero mean and unit variance)

\[
g(z|\eta, \lambda) = \begin{cases} bc(1 + \frac{1}{\eta - 2(1 - z)^2} -(\eta + 1)/2, & z < -a/b, \\ bc(1 + \frac{1}{\eta - 2(1 + z)^2} -(\eta + 1)/2, & z \geq -a/b, \end{cases}
\]

where \( 2 < \eta < \infty, -1 < \lambda < 1, \) and

\[
a = 4\lambda c \eta - 2 \eta - 1, \quad b = \sqrt{1 + 3\lambda^2 - a^2}, \quad c = \frac{\Gamma(\eta + 1/2)}{\sqrt{\pi(\eta - 2)\Gamma(\eta/2)}}.
\]
Following Hansen (1994), the following specification was adopted for the time-varying degrees of freedom:

\[ \eta_t = 2.1 + \frac{27.9}{1 + \exp(-\mu_t)}, \]

\[ \lambda_t = 0.9 \frac{1 - \exp(-v_t)}{1 + \exp(-v_t)}, \]

and the following estimations were obtained for the data of Section 6 (with estimated asymptotic standard errors in parentheses)

\[ \mu_t = -2.89 + 0.033 \varepsilon_{t-1}^+ + 0.013 \varepsilon_{t-1}^-, \]

\[ (0.42) (1.226) (0.450) \]

\[ v_t = 2.99 + 0.047 \varepsilon_{t-1}^+ + 0.149 \varepsilon_{t-1}^-, \]

\[ (0.16) (3.010) (20.31) \]

References


Lumsdaine, R.L., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64, 575–596.