DEPARTAMENTO DE ECONOMIA
PUC-RIO

TEXTO PARA DISCUSSÃO
Nº. 486

MODELING MULTIPLE REGIMES IN FINANCIAL VOLATILITY
WITH A FLEXIBLE COEFFICIENT GARCH MODEL

MARCELO C. MEDEIROS
ALVARO VEIGA

JULHO 2004

# MODELING MULTIPLE REGIMES IN FINANCIAL VOLATILITY WITH A FLEXIBLE COEFFICIENT GARCH(1,1) MODEL

MARCELO C. MEDEIROS AND ALVARO VEIGA

ABSTRACT. In this paper a flexible multiple regime GARCH(1,1)-type model is developed to describe the sign and size asymmetries and intermittent dynamics in financial volatility. The results of the paper are important to other nonlinear GARCH models. The proposed model nests some of the previous specifications found in the literature and has the following advantages: First, contrary to most of the previous models, more than two limiting regimes are possible and the number of regimes is determined by a simple sequence of of tests that circumvents identification problems that are usually found in nonlinear time series models. The second advantage is that the stationarity restriction on the parameters is relatively weak, thereby allowing for rich dynamics. It is shown that the model may have explosive regimes but can still be strictly stationary and ergodic. A simulation experiment shows that the proposed model can generate series with high kurtosis, low first-order autocorrelation of the squared observations, and exhibit the so-called "Taylor effect", even with Gaussian errors. Estimation of the parameters is addressed and the asymptotic properties of the quasi-maximum likelihood estimator are derived under weak conditions. A Monte-Carlo experiment is designed to evaluate the finite sample properties of the sequence of tests. Empirical examples are also considered.

KEYWORDS: Volatility, GARCH models, multiple regimes, nonlinear time series, smooth transition, finance, asymmetry, leverage effect, excess of kurtosis, asymptotic theory.

## 1. INTRODUCTION

MODELING AND FORECASTING the conditional variance, or volatility, of financial time series has been one of the major topics in financial econometrics. Forecasted conditional variances are used, for example, in portfolio selection, derivative pricing and hedging, risk management, market timing, and market making. Among solutions to tackle this problem, the ARCH (Autoregressive Conditional Heteroscedasticity) model proposed by Engle (1982) and the GARCH (Generalized Autoregressive Conditional Heteroscedasticity) specification introduced by Bollerslev (1986) are among the most widely used, and are now fully incorporated into financial econometric practice.

One drawback of the GARCH model is the symmetry in the response of volatility to past shocks, which fails to accommodate sign asymmetries. Starting with Black (1976), it has been observed that there is an asymmetric response of the conditional variance of the series to unexpected news, represented by shocks: Financial markets become more volatile in response to "bad news" (negative shocks) than to "good news" (positive shocks). Goetzmann, Ibbotson, and Peng (2001) found evidence of asymmetric sign effects in volatility as far back as 1857 for the NYSE. They report that unexpected negative shocks in the monthly return of the NYSE from 1857 to 1925 increase

volatility almost twice as much as equivalent positive shocks in returns. Similar results were also reported by Schwert (1990).

The above mentioned asymmetry has motivated a large number of different volatility models which have been applied with relatively success in several situations. Nelson (1991) proposed the Exponential GARCH (EGARCH) model. In his proposal, the natural logarithm of the conditional variance is modeled as a nonlinear ARMA model with a term that introduces asymmetry in the dynamics of the conditional variance, according to the sign of the lagged returns. Glosten, Jagannanthan, and Runkle (1993) proposed the GJR model, where the impact of the lagged squared returns on the current conditional variance changes according to the sign of the past return. A similar specification, known as Threshold GARCH (TGARCH), model was developed by Rabemananjara and Zakoian (1993) and Zakoian (1994). Ding, Granger, and Engle (1993) proposed the Asymmetric Power ARCH which nests several GARCH specifications. Engle and Ng (1993) popularized the news impact curve (NIC) as a measure of how new information is incorporated into volatility estimates. The authors also developed formal statistical tests to check the presence of asymmetry in the volatility dynamics. More recently, Fornari and Mele (1997) generalized the GJR model by allowing all the parameters to change according to the sign of the past return. Their proposal is known as the Volatility-Switching GARCH (VSGARCH) model. Based on the Smooth Transition AutoRegressive (STAR) model, Hagerud (1997) and Gonzalez-Rivera (1998) proposed the Smooth Transition GARCH (STGARCH) model. While the latter only considered the Logistic STGARCH (LSTGARCH) model, the former discussed both the Logistic and the Exponential STGARCH (ESTGARCH) alternatives. In the logistic STGARCH specification, the dynamics of the volatility are very similar to those of the GJR model and depends on the sign of the past returns. The difference is that the former allows for a smooth transition between regimes. In the EST-GARCH model, the sign of the past returns does not play any role in the dynamics of the conditional variance, but it is the magnitude of the lagged squared return that is the source of asymmetry. Anderson, Nam, and Vahid (1999) combined the ideas of Fornari and Mele (1997), Hagerud (1997), and Gonzalez-Rivera (1998) and proposed the Asymmetric Nonlinear Smooth Transition GARCH (ANSTGARCH) model, and found evidence in favor of their specification. Inspired by the Threshold Autoregressive (TAR) model, Li and Li (1996) proposed the Double Threshold ARCH (DTARCH) model. Liu, Li, and Li (1997) generalized it, proposing the Double Threshold GARCH (DT-GARCH) process to model both the conditional mean and the conditional variance as a threshold process. More recently, based on the regression-tree literature, Audrino and Bühlmann (2001) proposed the Tree Structured GARCH model to describe multiple limiting regimes in volatility [1].

In this paper we contribute to the literature by proposing a new flexible nonlinear GARCH model with multiple limiting regimes, called the Flexible Coefficient GARCH (FCGARCH) model, that nests several of the models mentioned above. As most of the empirical papers in the financial

[1]See also Cai (1994) and Hamilton and Susmel (1994) for regime switching GARCH specifications based on the Markov-switching model.

econometrics literature deal only with GARCH(1,1)-type of models, we focus our attention only on the first-order FCGARCH specification.

Our proposal has the following advantages: First, contrary to most of the previous models in the literature, more than two limiting regimes can be modeled. The number of regimes is determined by a simple and easily implemented sequence of tests that circumvents the identification problem in the nonlinear time series literature, and avoids the estimation of overfitted models. To the best of our knowledge, the only two exceptions that explicitly model more than two limiting regimes in the volatility are the DTGARCH and Tree-Structure GARCH models. However, in the former, the authors did not discuss how to determine the number of regimes and only one fixed threshold at zero is considered in the empirical application. In the latter, the proposed procedure is based on the use of information criteria and may suffer from identification problems when an irrelevant regime is estimated; see Hansen (1996) for a similar discussion considering threshold regression models and Teräsvirta and Mellin (1986) for the linear regression case. The second advantage is that the stationarity restriction on the FCGARCH model parameters is relatively weak, thereby allowing for rich dynamics. For example, the model may have explosive regimes and still be strictly stationary and ergodic, being capable of describing intermittent dynamics. The system spends a large fraction of time in a bounded region, but sporadically develops an instability that grows exponentially for some time, and then suddenly collapses. Furthermore, data with very high kurtosis can easily be generated even with Gaussian errors. This allows for a better description of the large absolute returns of financial time series that standard GARCH models fail to describe satisfactorily. Reproducing the above mentioned typical behavior of financial time series maybe important in risk analysis and management. A simulation experiment shows that the FCGARCH model is able to generate time series with high kurtosis and, at the same time, positive but low first-order autocorrelations of squared observations, which are frequently observed in financial time series. Furthermore, the FCGARCH model seems to be able to reproduce the so-called "Taylor effect" (Granger and Ding 1995). Other models such as the GARCH and the EGARCH models are not able to reproduce adequately the above mentioned stylized facts of financial time series; see Malmsten and Teräsvirta (2004) and Carnero, Peña, and Ruiz (2004) for comprehensive discussions.

We discuss the theoretical aspects of the FCGARCH model. Conditions for strict stationarity and for the existence of the second- and fourth-order moments; model identifiability; and the existence, consistency, and asymptotic normality of the quasi-maximum likelihood estimators. Consistency and asymptotic normality are proved under weak conditions. Our results are directly applicable to other nonlinear GARCH specifications, such as the STGARCH model. Furthermore, existing results in the literature are special cases of those presented in the paper.

A sequence of simple Lagrange multiplier (LM) tests is developed to determine the number of limiting regimes and to avoid the specification of models with an excessive number of parameters. Although the test is derived under the assumption that the errors are Gaussian, a robust version

against non-Gaussian errors is also considered. A Monte Carlo experiment is designed to evaluate
the finite sample properties of the proposed sequence of tests with simulated data. The main
finding is that the robust version of the test works well in small samples, and compares favorably
with the use of information criteria.

An empirical example with seven stock indexes shows evidence of two regimes for three series
and three regimes for other three series. Only for one stock index there is no evidence of regime
switching. Furthermore, for all series with three regimes, the GARCH model associated with
the first regime, representing very negative returns ("very bad news"), is explosive. The model
in the middle regime, related to tranquil periods, has a slightly lower persistence than the stan-
dard estimated GARCH(1,1) models in the literature. Finally, the third regime, representing large
positive returns, has an associated GARCH(1,1) specification that is significantly less persistent
than the others. Thus, we find strong evidence of both size and sign asymmetries. In addition,
the FCGARCH model produces normalized residuals with lower kurtosis than the GARCH and
GJR models. When a forecasting exercise is considered, the proposed model outperforms several
concurrent GARCH specifications.

The structure of the paper is as follows. Section 2 presents the model. Its probabilistic properties
are analyzed in Section 3. Estimation of the FCGARCH model is considered in Section 4. Section
5 discusses the test for an additional regime. Section 6 summarizes the modeling cycle procedure.
A Monte Carlo simulation is presented in Section 7, and empirical examples are considered in
Section 8. Finally, Section 9 concludes the paper. All technical proofs are given in the Appendix.

## 2. The Model

In this paper, we generalize the GARCH(1,1) and the Logistic STGARCH(1,1) formulations,
introducing a general regime switching scheme. The proposed model is defined as follows.

DEFINITION 1. *A time series* $\{y_t\}$ *follows a first-order Flexible Coefficient GARCH model with*
$m = H + 1$ *limiting regimes, FCGARCH$(m, 1, 1)$, if*

$$
y_t = h_t^{1/2} \varepsilon_t,
$$

$$
h_t = G\left(\mathbf{w}_t; \boldsymbol{\psi}\right) = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 +
$$

(1)

$$
\sum_{i=1}^{H} \left[\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2\right] f\left(s_t; \gamma_i, c_i\right), \quad t = 1, \ldots, T,
$$

*where* $\{\varepsilon_t\}$ *is a sequence of identically and independently distributed zero mean and unit variance
random variables,* $\varepsilon_t \sim \text{IID}(0, 1)$, $G\left(\mathbf{w}_t; \boldsymbol{\psi}\right)$ *is a nonlinear function of the vector of variables*
$\mathbf{w}_t = [y_{t-1}, h_{t-1}, s_t]'$, *and is indexed by the vector of parameters*

$$
\boldsymbol{\psi} = [\alpha_0, \beta_0, \lambda_0, \alpha_1, \ldots, \alpha_H, \beta_1, \ldots, \beta_H, \lambda_1, \ldots, \lambda_H, \gamma_1, \ldots, \gamma_H, c_1, \ldots, c_H,]' \in \mathbb{R}^{3+5H},
$$

and $f\left(s_{t};\gamma_{i},c_{i}\right)$, $i=1,\ldots,H$, *is the logistic function defined as*

$$(2) \qquad f\left(s_{t};\gamma_{i},c_{i}\right) = \frac{1}{1+e^{-\gamma_{i}(s_{t}-c_{i})}}.$$

It is clear that $f\left(s_{t};\gamma_{i},c_{i}\right)$ is a monotonically increasing function, such that $f\left(s_{t};\gamma_{i},c_{i}\right)\to1$ as $s_{t}\to\infty$ and $f\left(s_{t};\gamma_{i},c_{i}\right)\to0$ as $s_{t}\to-\infty$. The parameter $\gamma_{i}$, $i=1,\ldots,H$, is called the *slope parameter* and determines the speed of the transition between two limiting regimes. When $\gamma_{i}\to\infty$, the logistic function becomes a step function, and the FCGARCH model becomes a threshold-type specification. The variable $s_{t}$ is known as the *transition variable*. In this paper, we consider $s_{t}=y_{t-1}$. Hence, we model the differences in the dynamics of the conditional variance according to the sign and size of shocks in past returns, which represent previous "news". Of course, there are other possible choices for $s_{t}$; see Audrino and Trojani (forthcoming) and Chen, Chiang, and So (2003) for some alternatives.

The number of limiting regimes is defined by the hyper-parameter $H$. For example, suppose that in (1), $H=2$, $c_{1}$ is highly negative, and $c_{2}$ is very positive, than the resulting FCGARCH model will have 3 limiting regimes that can be interpreted as follows. The first regime may be related to extremely low negative shocks ("very bad news") and the dynamics of the volatility are driven by $h_{t}=\alpha_{0}+\beta_{0}h_{t-1}+\lambda_{0}y_{t-1}^{2}$ as $f\left(y_{t-1};\gamma_{i},c_{i}\right)\approx0$, $i=1,2$. In the the middle regime, which represents low absolute returns ("tranquil periods"), $h_{t}=\alpha_{0}+\alpha_{1}+(\beta_{0}+\beta_{1})h_{t-1}+(\lambda_{0}+\lambda_{1})y_{t-1}^{2}$ as $f\left(y_{t-1};\gamma_{1},c_{1}\right)\approx1$ and $f\left(y_{t-1};\gamma_{2},c_{2}\right)\approx0$. Finally, the third regime is related to high positive shocks ("very good news") and $h_{t}=\alpha_{0}+\alpha_{1}+\alpha_{2}+(\beta_{0}+\beta_{1}+\beta_{2})h_{t-1}+(\lambda_{0}+\lambda_{1}+\lambda_{2})y_{t-1}^{2}$, as $f\left(y_{t-1};\gamma_{i},c_{i}\right)\approx1$, $i=1,2$. As the speed of the transitions between different limiting GARCH models is determined by the parameter $\gamma_{i}$, $i=1,2$, the multiple regime interpretation of the FCGARCH specification will become clearer the more abrupt are the transitions ($\gamma_{i}\gg0$) [2]. In practical applications, the restriction $\gamma_{1}=\gamma_{2}=\cdots=\gamma_{H}$ may be imposed in order to reduce the number of parameters and the eventual computational cost of the estimation algorithm.

It is important to notice that model (1) nests several well-known GARCH specifications, such as:

- The GARCH(1,1) model if $\gamma_{i}=0$ or $\alpha_{i}=\beta_{i}=\lambda_{i}=0$, $i=1\ldots,h$.
- The LSTGARCH(1,1) model if $\alpha_{i}=\beta_{i}=0$, $i=1\ldots,h$ and $h=1$.
- The GJR(1,1) model if $H=1$, $\gamma_{1}\to\infty$, $\alpha_{1}=\beta_{1}=0$, and $c_{1}=0$.
- The VSGARCH(1,1) model if $H=1$ and $\gamma_{1}\to\infty$, $c_{1}=0$.
- The ANSTGARCH(1,1) model if $H=1$, and $c_{1}=0$.
- The variance component of the DTARCH(1,1) model if $\gamma_{i}\to\infty$ and $\alpha_{i}=\beta_{i}=0$, $i=1\ldots,h$.
- The variance component of the DTGARCH(1,1) model if $\gamma_{i}\to\infty$ and $s_{t}=h_{t-1}$.

---

[2] Representing multiple regimes with logistic functions dates back to Bacon and Watts (1971) and Chan and Tong (1986); see also Teräsvirta (1994) and van Dijk and Franses (1999).

The nonlinear GARCH model proposed in Lanne and Saikkonen (2005) is a special case of the FCGARCH model if $\beta_i = 0$, $i = 1, \ldots, H$, or if $s_t = h_{t-1}$. The FCGARCH model is a special case of the general GARCH specification presented in He and Teräsvirta (1999), Ling and McAleer (2002), and Carrasco and Chen (2002) if $s_t = \varepsilon_{t-1}$.

## 3. MAIN ASSUMPTIONS AND PROBABILISTIC PROPERTIES OF THE FCGARCH MODEL

We need to make the following set of assumptions:

ASSUMPTION 1. *The true parameter vector $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi} \subseteq \mathbb{R}^{3+5H}$ is in the interior of $\boldsymbol{\Psi}$, a compact and convex parameter space.*

ASSUMPTION 2. *The sequence $\{\varepsilon_t\}$ of $\mathsf{IID}(0,1)$ random variables is drawn from a continuous (with respect to Lebesgue measure on the real line), symmetric, unimodal, positive everywhere density, and bounded in a neighborhood of 0.*

ASSUMPTION 3. *The parameters $c_i$ and $\gamma_i$, $i = 1, \ldots, H$, satisfy the conditions:*
   (R.1) $-\infty < M < c_1 < \ldots < c_H < M < \infty$;
   (R.2) $\gamma_i > 0$.

ASSUMPTION 4. *The parameters $\gamma_i$ and $c_i$, $i = 1, \ldots, H$, are such that the logistic functions satisfy the following restrictions: $f(s_t; \gamma_1, c_1) \geq f(s_t; \gamma_2, c_2) \geq \ldots \geq f(s_t; \gamma_H, c_H)$, $\forall\, t \in [0, T]$.*

ASSUMPTION 5. *The parameters $\alpha_j$, $\beta_j$, and $\lambda_j$, $j = 0, \ldots, H$, satisfy the following restrictions:*
   (R.3) $\sum_{j=0}^{K} \alpha_j > 0$, $\forall K = 0, \ldots, H$;
   (R.3) $\sum_{j=0}^{K} \beta_j \geq 0$, and $\sum_{j=0}^{K} \lambda_j \geq 0$, $\forall K = 0, \ldots, H$.

Assumption 1 is standard. Assumption 2 is important for the mathematical derivations in this section and in Section 5. Assumption 3 guarantees the identifiability of the model (see Section 4.2 for details). The restrictions stated in Assumptions 4 and 5 ensure strictly positive conditional variances. Specifically, Assumption 4 ensures that the conditions in Assumption 5 are sufficient for the strict positivity of the conditional variance.

Defining $s_t = y_{t-1}$, model (1) may be written as

(3)
$$y_t = h_t^{1/2} \varepsilon_t,$$
$$h_t = g_{t-1} + c_{t-1} h_{t-1},$$

where

$$g_{t-1} \equiv g(y_{t-1}, \varepsilon_{t-1}) = \alpha_0 + \sum_{i=1}^{H} \alpha_i f_{i,t-1},$$

$$c_{t-1} \equiv c(y_{t-1}, \varepsilon_{t-1}) = \left[ \left( \beta_0 + \sum_{i=1}^{H} \beta_i f_{i,t-1} \right) + \left( \lambda_0 + \sum_{i=1}^{H} \lambda_i f_{i,t-1} \right) \varepsilon_{t-1}^2 \right],$$

with $f_{i,t-1} \equiv f\left(y_{t-1}; \gamma_i, c_i\right)$.

Following Nelson (1990), the next theorem states a necessary and sufficient log-moment condition for the strict stationarity and ergodicity of the FCGARCH($m$,1,1) model.

THEOREM 1. *Suppose that $y_t \in \mathbb{R}$ follows an FCGARCH(m,1,1) process as in (1), with $s_t = y_{t-1}$. Under Assumptions 2–5, the process $\mathbf{u}_t = (y_t, h_t)'$ is strictly stationary and ergodic if, and only if,*

$$
(4) \qquad \mathsf{E}\left\{\log\left[\left(\beta_0 + \sum_{i=1}^{H}\beta_i f_{i,t-1}\right) + \left(\lambda_0 + \sum_{i=1}^{H}\lambda_i f_{i,t-1}\right)\varepsilon_{t-1}^2\right]\right\} < 0, \, \forall\, t.
$$

*Furthermore, there is a second-order stationary solution to (3) that has the following causal expansion:*

$$
y_t = h_t^{1/2}\varepsilon_t,
$$

$$
(5)
$$

$$
h_t = g_{t-1} + \sum_{k=0}^{\infty}\prod_{j=0}^{k} g_{t-1-k}c_{t-1-j},
$$

*where the infinite sum converges almost surely (a.s.).*

The log-moment condition is important as the condition in Theorem 1 can be satisfied even in the absence of finite second-moments of $y_t$; see McAleer (2005) for a comprehensive discussion of log-moment conditions for volatility models.

COROLLARY 1. *Under the assumptions of Theorem 1, a sufficient condition for strict stationary and ergodicity of $\mathbf{u}_t = (y_t, h_t)'$ in terms of the parameters is*

$$
\frac{1}{2}\left(\beta_0 + \lambda_0\right) + \frac{1}{2}\sum_{i=0}^{H}\left(\beta_i + \lambda_i\right) \le 1.
$$

Deriving a general sufficient condition for the existence of the moments of $y_t$ is rather complicated. However, the moment condition stated in the following theorem can be used to find a necessary and sufficient condition for the existence of low-order moments of $y_t$. As mentioned in the previous section, the model families of He and Teräsvirta (1999), Ling and McAleer (2002), Lanne and Saikkonen (2005), and Carrasco and Chen (2002) do not nest the FCGARCH model without additional restrictions. Hence, the direct application of the results of these authors is not straightforward. In the subsequent corollary, we derive sufficient conditions for the existence of the second- and forth-order moments of $y_t$.

THEOREM 2. *Suppose that $y_t \in \mathbb{R}$ follows an FCGARCH(m,1,1) process as in (1), with $s_t = y_{t-1}$ and $\mathsf{E}\left[\varepsilon_t^{2k}\right] = \mu_{2k} < \infty$, for $k = 1, 2, 3, \ldots.$. Under Assumptions 2–5, and assuming that the*

*moments of order up to $n = k - 1$ exist, $\mathsf{E}\left[y_t^{2n}\right] < \infty$, the $2k$th-order moment of $y_t$ exists if*

$$(6) \qquad \mathsf{E}\left\{\left[\beta_0 + \lambda_0\varepsilon_{t-1}^2 + \sum_{i=1}^{H}\left(\beta_i + \lambda_i\varepsilon_{t-1}^2\right)f_{i,t-1}\right]^k\right\} < 1.$$

COROLLARY 2. *Suppose that $y_t \in \mathbb{R}$ follows an FCGARCH(m,1,1) process as in (1), with $s_t = y_{t-1}$. A sufficient condition for the existence of the second-order moment of $y_t$ is*

$$(7) \qquad \frac{1}{2}\left(\beta_0 + \lambda_0\right) + \frac{1}{2}\sum_{i=0}^{H}\left(\beta_i + \lambda_i\right) < 1.$$

*Furthermore, define $\beta_U \equiv \sum_{i=1}^{H}\beta_i$ and $\lambda_U \equiv \sum_{i=1}^{H}\lambda_i$. Under Assumptions 2–5, the fourth-order moment of $y_t$ exists if $\mathsf{E}\left[\varepsilon_t^4\right] = \mu_4 < \infty$, (7) holds, and*

$$(8) \quad \beta_0^2 + \beta_0\beta_U + \frac{\beta_U^2}{2} + \mu_4\left[\lambda_0 + \lambda_0\lambda_U + \frac{\lambda_U^2}{2}\right] + 2\lambda_0\beta_0 + \beta_0\lambda_U + \lambda_0\beta_U + \lambda_U\beta_U < 1.$$

REMARK 1. *When $H = 0$, conditions (7) and (8) are the usual conditions for the existence of the second- and fourth-order moments of GARCH models. When $H = 1$, $\gamma_1 \to \infty$, $\alpha_1 = 0$, and $\beta_1 = 0$, conditions (7) and (8) become the usual ones for the GJR model.*

It is important to notice that, even with explosive regimes the FCGARCH$(m, 1, 1)$ may still be strictly stationary, ergodic, and with finite fourth-order moment. Furthermore, some of the parameters of the limiting GARCH models may exceed one. This flexibility generates models with higher kurtosis than the standard GARCH(1,1), even with Gaussian errors.

REMARK 2. *The IGARCH model with Gaussian errors is also capable of generating data with high kurtosis. However, contrary to the FCGARCH model, it does not have finite second- and fourth-order moments.*

The following examples illustrate some interesting situations.

Consider 3000 replications of the following FCGARCH$(3, 1, 1)$ models with Gaussian errors, each of which has 5000 observations.

(1) Example 1:

$$y_t = h_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathsf{NID}(0,1)$$
$$h_t = 1 \times 10^{-4} + 0.96h_{t-1} + 0.18y_{t-1}^2 +$$
$$\left(-0.9 \times 10^{-4} - 0.60h_{t-1} - 0.10y_{t-1}^2\right)f\left(5000\left(y_{t-1} + 0.005\right)\right) +$$
$$\left(1 \times 10^{-4} + 0.10h_{t-1} + 0.05y_{t-1}^2\right)f\left(5000\left(y_{t-1} - 0.02\right)\right).$$

(2) Example 2:

$$y_t = h_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathsf{NID}(0,1)$$

$$h_t = 6 \times 10^{-5} + 1.10 h_{t-1} + 0.10 y_{t-1}^2 +$$

$$\left(-5 \times 10^{-5} - 0.65 h_{t-1} - 0.09 y_{t-1}^2\right) f\left(3000\left(y_{t-1} + 0.005\right)\right) +$$

$$\left(1 \times 10^{-5} + 0.10 h_{t-1} + 0.04 y_{t-1}^2\right) f\left(3000\left(y_{t-1} - 0.005\right)\right).$$

(3) Example 3:

$$y_t = h_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathsf{NID}(0,1)$$

$$h_t = 6 \times 10^{-5} + 1.20 h_{t-1} + 0.10 y_{t-1}^2 +$$

$$\left(-5.5 \times 10^{-5} - 1.20 h_{t-1} - 0.10 y_{t-1}^2\right) f\left(2000\left(y_{t-1} + 0.001\right)\right) +$$

$$5 \times 10^{-5} f\left(2000\left(y_{t-1} - 0.01\right)\right).$$

The models in Examples 1–3 have three extreme regimes, each with the first regime being explosive as $\beta_0 + \lambda_0 > 1$. However, even with an explosive regime, the generated time series are still stationary provided that $\frac{1}{2}\left(\beta_0 + \lambda_0\right) + \frac{1}{2}\sum_{i=0}^{2}\left(\beta_i + \lambda_i\right) < 1$. Furthermore, the fourth-order moment exists, provided that condition (8) is also satisfied. Note also that $\beta_0 > 1$ in Examples 2 and 3. The model in Example 3 has the interesting property that the GARCH effect is only present in the extreme regimes. The regime associated with tranquil periods is homoskedastic.

Figure 1 shows the scatter plot of the estimated kurtosis and first-order autocorrelation of the squared observations. The dots indicates the cases where the first-order autocorrelation of $|y_t|$ is greater than the first-order autocorrelation of $|y_t|^2$. The crosses indicate the opposite effect. The simulated FCGARCH models seem to reproduce some of the stylized facts observed in financial time series. Table 1 summarizes some statistics about the estimated kurtosis and autocorrelations. As can be seen, the minimum value of the estimated kurtosis is over 3. In addition the mean values of the estimated first-order autocorrelations are in accordance with the typical numbers that are found in practical applications.

TABLE 1. SIMULATED MODELS: DESCRIPTIVE STATISTICS.

The table shows descriptive statistics for the estimated kurtosis and first-order autocorrelation of the squared observations over 3000 replications of Models (1)–(3).

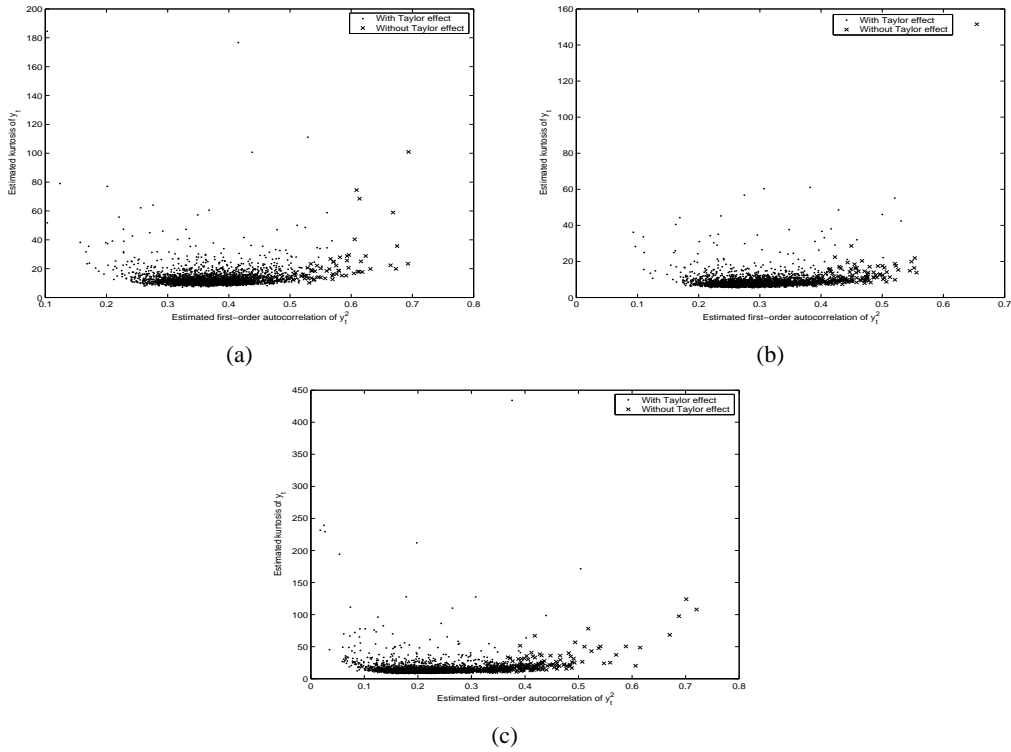| | Kurtosis | | | | Autocorrelation | | | |
|---|---|---|---|---|---|---|---|---|
| Example | Min. | Max. | Mean | Std. Dev. | Min. | Max. | Mean | Std. Dev. |
| 1 | 7.18 | 184.51 | 13.42 | 7.72 | 0.10 | 0.69 | 0.37 | 0.07 |
| 2 | 5.37 | 151.50 | 8.81 | 4.75 | 0.09 | 0.65 | 0.29 | 0.06 |
| 3 | 7.75 | 434.12 | 15.88 | 14.99 | 0.02 | 0.72 | 0.22 | 0.08 |

FIGURE 1. Scatter plot of the estimated kurtosis and first-order autocorrelation of the squared observations. The dots indicates the cases where the "Taylor effect" is satisfied and the crosses indicate the opposite effect. Panel (a) concerns Example 1. Panel (b) concerns Example 2. Panel (c) concerns Example 3.

## 4. PARAMETER ESTIMATION

As the distribution of $\varepsilon_t$ is unknown, the parameters of the FCGARCH model are estimated by quasi-maximum likelihood (QML). For GARCH(1,1) models, Lee and Hansen (1994) proved that the local QMLE is consistent and asymptotically normal if all the conditional expectations of $\varepsilon_t^{2+\kappa} < \infty$ uniformly with $\kappa > 0$. Lumsdaine (1996) required that $\mathsf{E}\left[\varepsilon_t^{32}\right] < \infty$. Jeantheau (1998) discussed consistency of the QMLE under weaker conditions. More recently, Ling and McAleer (2003) proved the consistency of the global QMLE for a VARMA-GARCH model under only the second-order moment condition. The authors also proved the asymptotic normality of the global (local) QMLE under the sixth-order (forth-order) moment condition. Comte and Lieberman (2003) and Berkes, Horváth, and Kokoszka (2003) proved consistency and asymptotic normality of the QMLE of the parameters of the GARCH($p,q$) model under the second- and fourth-order moment conditions, respectively.

As in Boussama (2000), McAleer, Chan, and Marinova (forthcoming), and Francq and Zakoïan (2004), we prove consistency and asymptotic normality of the QMLE of the FCGARCH($m, 1, 1$) under the log-moment condition in Theorem 1; see also Li, Ling, and McAleer (2002) and McAleer

(2005). Extending the results in Jensen and Rahbek (2004) for non-stationary ARCH models to the case of the FCGARCH model is not straightforward, and is beyond the scope of this paper. However, this is an interesting topic for future research.

The quasi-log-likelihood function of the FCGARCH model is given by

$$(9) \qquad \mathcal{L}_T(\boldsymbol{\psi}) = \frac{1}{T} \sum_{t=1}^{T} l_t(\boldsymbol{\psi}),$$

where $l_t(\boldsymbol{\psi}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_t) - \frac{y_t^2}{2h_t}$. Note that the processes $y_t$ and $h_t$, $t \leq 0$, are unobserved, and hence they are arbitrary constants. Thus, $\mathcal{L}_T(\boldsymbol{\psi})$ is a quasi-log-likelihood function that is not conditional on the true $(y_0, h_0)$ making it suitable for practical applications.

However, to prove the asymptotic properties of the QMLE is more convenient to work with the unobserved process $\{(y_{u,t}, h_{u,t}) : t = 0, \pm 1, \pm 2, \ldots\}$, which satisfies

$$(10) \qquad h_{u,t} = \alpha_0 + \beta_0 h_{u,t-1} + \lambda_0 y_{u,t-1}^2 + \sum_{i=1}^{H} \left[ \alpha_i + \beta_i h_{u,t-1} + \lambda_i y_{u,t-1}^2 \right] f\left( y_{u,t-1}; \gamma_i, c_i \right).$$

The unobserved quasi-log-likelihood function conditional on $\mathcal{F}_0 = (y_0, y_{-1}, y_{-2}, \ldots)$ is

$$(11) \qquad \mathcal{L}_{u,T}(\boldsymbol{\psi}) = \frac{1}{T} \sum_{t=1}^{T} l_{u,t}(\boldsymbol{\psi}),$$

with $l_{u,t}(\boldsymbol{\psi}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_{u,t}) - \frac{y_{u,t}^2}{2h_{u,t}}$. The primary difference between $\mathcal{L}_T(\boldsymbol{\psi})$ and $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is that the former is conditional on any initial values, whereas the latter is conditional on an infinite series of past observations. In practical situations, the use of (11) is not possible.

Let

$$\widehat{\boldsymbol{\psi}}_T = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\operatorname{argmax}} \mathcal{L}_T(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\operatorname{argmax}} \left( \frac{1}{T} \sum_{t=1}^{T} l_t(\boldsymbol{\psi}) \right),$$

and

$$\widehat{\boldsymbol{\psi}}_{u,T} = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\operatorname{argmax}} \mathcal{L}_{u,T}(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\operatorname{argmax}} \left( \frac{1}{T} \sum_{t=1}^{T} l_{u,t}(\boldsymbol{\psi}) \right).$$

Define $\mathcal{L}(\boldsymbol{\psi}) = \mathsf{E}\left[l_{u,t}(\boldsymbol{\psi})\right]$. In the following two subsections, we discuss the existence of $\mathcal{L}(\boldsymbol{\psi})$ and the identifiability of the FCGARCH model. Then, in Subsection 4.3, we prove the consistency of $\widehat{\boldsymbol{\psi}}_T$ and $\widehat{\boldsymbol{\psi}}_{u,T}$. We first prove the consistency of $\widehat{\boldsymbol{\psi}}_{u,T}$. Using Lemma 3 in Appendix B, we show that $\underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\sup} |\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})| \xrightarrow{p} 0$, and the consistency of $\widehat{\boldsymbol{\psi}}_T$ follows. The asymptotic normality of both estimators is considered in Subsection 4.4. We start proving asymptotic normality of $\widehat{\boldsymbol{\psi}}_{u,T}$. Then, using the results of Lemma 5, the proof for $\widehat{\boldsymbol{\psi}}_T$ is straightforward.

4.1. **Existence of the QMLE.** The following theorem proves the existence of $\mathcal{L}(\boldsymbol{\psi})$. It is based on Theorem 2.12 in White (1994), which establishes that, under certain conditions of continuity and measurability on quasi-log-likelihood function, $\mathcal{L}(\boldsymbol{\psi})$ exists.

THEOREM 3. *If (4) is satisfied, under Assumptions 2–5, $\mathcal{L}(\boldsymbol{\psi})$ exists and is finite.*

4.2. **Identifiability of the Model.** A fundamental problem for statistical inference with nonlinear time series models is the unidentifiability of the parameters. In order to guarantee unique identifiability of the quasi-log-likelihood function, the sources of uniqueness of the model must be examined. Here, the main concepts and results will be discussed briefly. In particular, the conditions that guarantee that the FCGARCH model is identifiable and minimal will be established and proved. First, two related concepts will be discussed: The concept of minimality of the model, established in Sussman (1992), also called "non-redundancy" in Hwang and Ding (1997); and the concept of reducibility of the model.

DEFINITION 2. *The FCGARCH$(m, 1, 1)$ model is minimal (or non-redundant) if its input-output map cannot be obtained from an FCGARCH$(n, 1, 1)$ model, where $n < m$.*

One source of unidentifiability comes from the fact that a model may contain irrelevant "limiting regimes". A limiting regime is represented by the functions

$$\mu_i = \left[ \alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2 \right] f(y_{t-1}; \gamma_i, c_i), \, i = 1, \ldots, H.$$

This means that there are cases where the model can be reduced without changing the input-output map. Thus, the minimality condition can only hold for irreducible models.

DEFINITION 3. *Define $\boldsymbol{\theta}_i = [\gamma_i, c_i]'$ and let $\varphi(y_{t-1}; \boldsymbol{\theta}_i) = \gamma_i(y_{t-1} - c_i)$, $i = 1, \ldots, H$. The FCGARCH model defined in (1) is reducible if one of the following three conditions holds:*

  (1) *One of the triples $(\alpha_i, \beta_i, \lambda_i)$ vanishes jointly for some $i \in [1, H]$;*
  (2) *$\gamma_i = 0$ for some $i \in [1, H]$;*
  (3) *There is at least one pair $(i, j)$, $i \neq j$, $i = 1, \ldots, H$, $j = 1, \ldots, H$, such that $|\varphi(y_{t-1}; \boldsymbol{\theta}_i)| = |\varphi(y_{t-1}; \boldsymbol{\theta}_j)|$, $\forall y_{t-1} \in \mathbb{R}$, $t = 1, \ldots, T$ (sign-equivalence).*

DEFINITION 4. *The FCGARCH model is identifiable if there are no two sets of parameters such that the corresponding distributions of the population variable $y$ are identical.*

Three properties of the FCGARCH model cause unidentifiability of the models:

  (P.1) The property of interchangeability of the regimes. The value of the likelihood function of the model does not change if the regimes are permuted. This results in $H!$ different models that are indistinct among themselves. As a consequence, in the estimation of the parameters, we will have $H!$ equal local maxima for the quasi-log-likelihood function.
  (P.2) The fact that $f(y_{t-1}; \gamma_i, c_i) = 1 - f(y_{t-1}; -\gamma_i, c_i)$.
  (P.3) Conditions $(1) - (2)$ in the definition of reducibility provide information about the presence of irrelevant regimes, which translate into identifiability sources. If the model contains a regime such that $\alpha_i = 0$, $\beta_i = 0$, and $\lambda_i = 0$, then the parameters $\gamma_i$ and $c_i$ remain unidentified, for some $i \in [1, H]$. On the other hand, if $\gamma_i = 0$, then the parameters $\alpha_i$, $\beta_i$, $\lambda_i$, and $c_i$ may take on any value without changing the quasi-log-likelihood function.

Property (P.3) is related to the concept of reducibility. In the same spirit of the results stated in Sussman (1992) and Hwang and Ding (1997), we show that, if the model is irreducible, properties (P.1) and (P.2) are the only forms of modifying the parameters without affecting the log-likelihood. Hence, by establishing the restrictions on the parameters of (1) that simultaneously avoid model reducibility, any permutation of regimes, and symmetries in the logistic function, we guarantee the identifiability of the model.

The problem of interchangeability, (P.1), can be prevented by imposing the Restrictions (R.1) in Assumption 3. The consequences due to the symmetry of the logistic function (P.2) can be resolved if we consider Restrictions (R.2) in Assumption 3. The presence of irrelevant regimes, (P.3), can be circumvented by applying a "specific-to-general" modeling strategy as will be suggested in Section 5.

Corollary 2.1 in Sussman (1992) and Corollary 2.4 in Hwang and Ding (1997) guarantee that an irreducible model is minimal. The fact that irreducibility and minimality are equivalent implies that there are no mechanisms, other than those listed in the definition of irreducibility, that can be used to reduce the complexity of the model without changing the functional input-output relation. Then, the restrictions in Assumption 3 guarantee that, if irrelevant regimes do not exist the model is identifiable and minimal.

We need an additional assumption before establishing the sufficient conditions under which the FCGARCH model is globally identifiable.

ASSUMPTION 6. *The parameters $\alpha_i$, $\beta_i$, and $\lambda_i$ do not vanish jointly for some $i \in [1, H]$.*

Assumption 6 guarantees that there are no irrelevant regimes.

THEOREM 4. *Under Assumptions 3 and 6, the FCGARCH$(m, 1, 1)$ model is globally identifiable. Furthermore, $\mathcal{L}(\boldsymbol{\psi})$ is uniquely maximized at $\boldsymbol{\psi}_0$.*

4.3. **Consistency.** The proof of consistency of the QMLE for the FCGARCH model follows the same reasoning given in Ling and McAleer (2003). The following theorem states and proves the main consistency result.

THEOREM 5. *If (4) is satisfied, under Assumptions 1–5, $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{p} \boldsymbol{\psi}_0$ and $\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_0$.*

4.4. **Asymptotic Normality.** In order to prove asymptotic normality, we define:

$$\mathbf{A}(\boldsymbol{\psi}_0) = \mathsf{E}\left[ -\frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\bigg|_{\boldsymbol{\psi}_0} \right] \text{ and}$$

$$\mathbf{B}(\boldsymbol{\psi}_0) = \mathsf{E}\left[ T \frac{\partial \mathcal{L}_{u,T}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}_0} \frac{\partial \mathcal{L}_{u,T}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'}\bigg|_{\boldsymbol{\psi}_0} \right] \equiv \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}\left( \frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}_0} \frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'}\bigg|_{\boldsymbol{\psi}_0} \right).$$

Consider the additional matrices:

$$\mathbf{A}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\psi}}\frac{\partial h_t}{\partial \boldsymbol{\psi}'}\left(\frac{y_t^2}{h_t}\right) - \left(\frac{y_t^2}{h_t} - 1\right)\frac{\partial}{\partial \boldsymbol{\psi}'}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \boldsymbol{\psi}}\right)\right\}, \text{ and}$$

(12)

$$\mathbf{B}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'} = \frac{1}{4T}\sum_{t=1}^{T}\frac{1}{h_t^2}\left(\frac{y_t^2}{h_t} - 1\right)^2\frac{\partial h_t}{\partial \boldsymbol{\psi}}\frac{\partial h_t}{\partial \boldsymbol{\psi}'}.$$

The following theorem states the asymptotic normality result.

THEOREM 6. *If (4) is satisfied and* $\mathsf{E}\left[\varepsilon_t^4\right] = \mu_4 < \infty$, *under Assumptions 1–5,*

(13)
$$T^{1/2}(\widehat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}_0) \xrightarrow{D} \mathsf{N}\left(\mathbf{0}, \mathbf{A}(\boldsymbol{\psi}_0)^{-1}\mathbf{B}(\boldsymbol{\psi}_0)\mathbf{A}(\boldsymbol{\psi}_0)^{-1}\right),$$

*where* $\mathbf{A}(\boldsymbol{\psi}_0)$ *and* $\mathbf{B}(\boldsymbol{\psi}_0)$ *are consistently estimated by* $\mathbf{A}_T(\widehat{\boldsymbol{\psi}})$ *and* $\mathbf{B}_T(\widehat{\boldsymbol{\psi}})$, *respectively.*

REMARK 3. *Under Assumption 2, it is clear that* $\mathbf{B}(\boldsymbol{\psi}_0) = \frac{1}{2}\left(\mathsf{E}\left[\varepsilon_t^4\right] - 1\right)\mathbf{A}(\boldsymbol{\psi}_0)$, *which reduces to the information matrix equality when* $\mathsf{E}\left[\varepsilon_t^4\right] = 3$.

## 5. DETERMINING THE NUMBER OF REGIMES

The number of regimes in the FCGARCH model, as represented by the number of transition functions in (1), is not known in advance and should be determined from the data. One possibility is to begin with a small model (such as GARCH(1,1) or white noise) and add regimes sequentially. The decision to add another regime may be based on the use of model selection criteria (MSC) or cross-validation. For example, Audrino and Bühlmann (2001) used Akaike's Information Criterion (AIC) to select the number of regimes in their Tree-Structured GARCH model. However, this has the following drawback. Suppose the data have been generated by an FCGARCH model with $m$ regimes ($m - 1$ transition functions). Applying MSC to decide whether or not another regime should be added to the model requires estimation of a model with $m$ logistic functions. In this situation, the larger model is not identified and its parameters cannot be estimated consistently [3]. This is likely to cause numerical problems in quasi-maximum likelihood estimation. Even when convergence is achieved, lack of identification causes a severe problem in interpreting the MSC. The FCGARCH model with $m$ regimes is nested in the model with $m + 1$ regimes.

A typical MSC comparison of the two models is then equivalent to a likelihood ratio test of $m$ against $m + 1$ regimes; see Teräsvirta and Mellin (1986) for a discussion. The choice of MSC determines the (asymptotic) significance level of the test. When the larger model is not identified under the null hypothesis, the likelihood ratio statistic does not have an asymptotic $\chi^2$ distribution under the null.

---

[3]In the case of the tree-structured GARCH model of Audrino and Bühlmann (2001), the identification issue is related to the location of the threshold. When an irrelevant regime is added, the location of the split cannot be estimated consistently; see Hansen (1996) for a discussion.

In this paper we tackle the problem of determining the number of regimes of the FCGARCH model with a "specific-to-general" modeling strategy, but we circumvent the problem of identification in a way that enables us to control the significance level of the tests in the sequence, and compute an upper bound to the overall significance level [4].

The following is based on the assumption that the errors $\varepsilon_t$ are Gaussian, but the results will be made robust to nonnormal errors.

Consider an FCGARCH with $H$ limiting regimes, defined as

$$
y_t = h_t^{1/2} \varepsilon_t,
$$

(14)
$$
h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \sum_{i=1}^{H-1} \left[ \alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2 \right] f(y_{t-1}; \gamma_i, c_i).
$$

The idea is to test the presence of an additional regime, as represented by an extra term in (14) of the form

(15)
$$
\left[ \alpha_H + \beta_H h_{t-1} + \lambda_H y_{t-1}^2 \right] f(y_{t-1}; \gamma_H, c_H)
$$

A convenient null hypothesis is

(16)
$$
\mathcal{H}_0 : \gamma_H = 0,
$$

against the alternative $\mathcal{H}_a : \gamma_H > 0$. Note that model (14) is not identified under the null hypothesis. In order to remedy this problem, we follow Lundbergh and Teräsvirta (2002) and expand the logistic function $f(y_{t-1}; \gamma_H, c_H)$ into a first-order Taylor expansion around the null hypothesis $\gamma_H = 0$ [5]. After merging terms, the resulting model for $h_t$ is

(17)
$$
h_t = \tilde{\alpha}_0 + \tilde{\beta}_0 h_{t-1} + \tilde{\lambda}_0 y_{t-1}^2 + \sum_{i=1}^{H-1} \left[ \alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2 \right] f(y_{t-1}; \gamma_i, c_i) +
$$
$$
\pi y_{t-1} + \delta h_{t-1} y_{t-1} + \rho y_{t-1}^3 + R,
$$

where $R$ is the remainder, $\tilde{\alpha}_0 = \alpha_0 - \frac{\alpha_H \gamma_H c_H}{4}$, $\tilde{\beta}_0 = \beta_0 - \frac{\beta_H \gamma_H c_H}{4}$, $\tilde{\lambda}_0 = \lambda_0 - \frac{\lambda_H \gamma_H c_H}{4}$, $\pi = \frac{\gamma_H}{4}$, $\delta = \frac{\beta_H \gamma_H}{4}$, and $\rho = \frac{\lambda_H \gamma_H}{4}$.

Define $f_{i,t-1} \equiv f(y_{t-1}; \gamma_i, c_i)$, $i = 1, \ldots, H$. Under $\mathcal{H}_0$, $R = 0$ and the quasi-maximum likelihood approach enables us to state the following result:

THEOREM 7. *If the stationarity condition in Theorem 1 is satisfied, under Assumptions 2–5 and the additional assumption that* $\mathsf{E}\left[ \left| y_t^6 \right| \right] < \infty$ *under the null, the LM statistic given by*

(18)
$$
LM = \frac{T}{2} \left\{ \sum_{t=1}^{T} \left( \frac{y_t^2}{\widehat{h}_{0,t}} - 1 \right) \widehat{\mathbf{d}}_t \right\} \left[ \sum_{t=1}^{T} \widehat{\mathbf{d}}_t \widehat{\mathbf{d}}_t' \right]^{-1} \left\{ \sum_{t=1}^{T} \left( \frac{y_t^2}{\widehat{h}_{0,t}} - 1 \right) \widehat{\mathbf{d}}_t \right\},
$$

---

[4] An equivalent procedure has been adopted in Medeiros and Veiga (2005) and Medeiros, Teräsvirta, and Rech (in press).

[5] The idea of circumventing the identification problem by approximating the nonlinear contribution by a low-order Taylor expansion under the null was originally proposed by Luukkonen, Saikkonen, and Teräsvirta (1988).

*where $\widehat{h}_{0,t}$ is the estimated conditional variance of the process under the null, $\widehat{\mathbf{d}}_t = [\widehat{\mathbf{z}}'_t, \widehat{\mathbf{u}}'_t]'$,*

$$\widehat{\mathbf{z}}_t = \frac{1}{\widehat{h}_{0,t}} \left.\frac{\partial h_t}{\partial \boldsymbol{\psi}'}\right|_{\mathcal{H}_0} = \frac{1}{\widehat{h}_{0,t}} \left\{ \widehat{\mathbf{x}}_t + \sum_{k=1}^{t-1} \left[ \prod_{j=k+1}^{t} \left( \widehat{\beta}_0 + \sum_{i=1}^{H-1} \widehat{\beta}_i f_{i,j-1} \right) \right] \widehat{\mathbf{x}}_k \right\},$$

$$\widehat{\mathbf{u}}_t = \frac{1}{\widehat{h}_{0,t}} \left.\frac{\partial h_t}{\partial \boldsymbol{\theta}'}\right|_{\mathcal{H}_0} = \frac{1}{\widehat{h}_{0,t}} \left\{ \widehat{\mathbf{v}}_t + \sum_{k=1}^{t-1} \left[ \prod_{j=k+1}^{t} \left( \widehat{\beta}_0 + \sum_{i=1}^{H-1} \widehat{\beta}_i f_{i,j-1} \right) \right] \widehat{\mathbf{v}}_k \right\},$$

$$\widehat{\mathbf{x}}_t = \Big[ 1, \widehat{h}_{0,t-1}, y^2_{t-1}, f_{1,t-1}, \ldots, f_{H-1,t-1},$$

$$f_{1,t-1}\widehat{h}_{0,t-1}, \ldots, f_{H-1,t-1}\widehat{h}_{0,t-1}, f_{1,t-1}y^2_{t-1}, \ldots, f_{H-1,t-1}y^2_{t-1},$$

$$(\widehat{\alpha}_1 + \widehat{\beta}_1 h_{0,t-1} + \widehat{\lambda}_1 y^2_{t-1})\frac{\partial f_{1,t-1}}{\partial \gamma_1}, \ldots, (\widehat{\alpha}_{H-1} + \widehat{\beta}_{H-1}\widehat{h}_{0,t-1} + \widehat{\lambda}_{H-1}y^2_{t-1})\frac{\partial f_{H-1,t-1}}{\partial \gamma_{H-1}},$$

$$(\widehat{\alpha}_1 + \widehat{\beta}_1 h_{0,t-1} + \widehat{\lambda}_1 y^2_{t-1})\frac{\partial f_{1,t-1}}{\partial c_1}, \ldots, (\widehat{\alpha}_{H-1} + \widehat{\beta}_{H-1} h_{0,t-1} + \widehat{\lambda}_{H-1}y^2_{t-1})\frac{f_{H-1,t-1}}{\partial c_{H-1}} \Big]',$$

$$\widehat{\mathbf{v}}_t = \Big[ y_{t-1}, \widehat{h}_{0,t-1}y_{t-1}, y^3_{t-1} \Big]',$$

*and*

$$\frac{\partial f_{i,t-1}}{\partial \gamma_i} = f_{i,t-1}(1 - f_{i,t-1})(y_{t-1} - c_i), \ i = 1, \ldots, H,$$

$$\frac{\partial f_{i,t-1}}{\partial c_i} = -f_{i,t-1}(1 - f_{i,t-1})\gamma_i, \ i = 1, \ldots, H,$$

*has a $\chi^2$ distribution with 3 degrees of freedom under the null hypothesis.*

REMARK 4. *The sixth-order moment condition is necessary for the existence of $\mathsf{E}[\mathbf{v}_t \mathbf{v}'_t]$.*

Under the normality assumption, the test can be performed in stages, as follows.

(1) Estimate model (1) under the null, call the estimated variance $\widehat{h}_{0,t}$, and compute $SSR_0 = \sum_{t=1}^{T} \left( y^2_t / \widehat{h}_{0,t} - 1 \right)^2$.

(2) Regress $\left( y^2_t / \widehat{h}_{0,t} - 1 \right)$ on $\widehat{\mathbf{z}}_t$ and $\widehat{\mathbf{u}}_t$ and compute the sum of the squared residuals, $SSR_1$.

(3) Compute the LM statistic

$$(19) \qquad LM = T\frac{SSR_0 - SSR_1}{SSR_0},$$

or the F statistic

$$(20) \qquad F = \frac{(SSR_0 - SSR_1)/3}{SSR_1/(T - 5H + 2)}.$$

Under $\mathcal{H}_0$, $LM$ is approximately distributed as $\chi^2$ with $p$ degrees of freedom and $F$ has an F distribution with 3 and $T - 5H + 2$ degrees of freedom.

Although the test statistic is constructed under the assumption of normality, we can follow Lundbergh and Teräsvirta (2002) and consider a robust version of the LM test against nonnormal errors. The robust version of the test can be constructed following the Procedure 4.1 of Wooldridge (1990). The test is performed as follows:

(1) As above.
(2) Regress $\widehat{\mathbf{u}}_t$ on $\widehat{\mathbf{z}}_t$ and compute the residual vectors, $\widehat{\mathbf{r}}_t$, $t = 1, \ldots, T$.
(3) Regress 1 on $\left(\frac{y_t^2}{\widehat{h}_{0,t}} - 1\right) \widehat{\mathbf{r}}_t$, and compute the residual sum of squares, SSR. The test statistic given by

$$(21) \qquad LM_R = T - SSR$$

has an asymptotic $\chi^2$ distribution with 3 degrees of freedom under the null hypothesis.

As observed in Lundbergh and Teräsvirta (2002), the robust version of the LM test should always be preferred to the nonrobust tests. At relevant sample sizes when the errors are normal, they are about as powerful as the normality-based LM tests.

Finally, it is important to stress that the results of the sequence of LM tests may be affected by possible outliers in the data. Nevertheless, an outlier-robust version of the LM test can be easily developed, following van Dijk, Franses, and Lucas (1999a,1999b).

## 6. MODELING CYCLE

We are now ready to combine the above statistical ingredients into a practical modeling strategy. We begin by testing linearity against an ARCH($q$) model at significance level $\delta$ [6]. The model under the null hypothesis is an homoskedastic model. If the null hypothesis is not rejected, the homoskedastic model is considered as the data generating process. In case of rejection, a GARCH(1,1) model is estimated and tested against an FCGARCH(1,1,1) model with two regimes at the significance level $\delta\varrho$, $0 < \varrho < 1$. Another rejection leads to estimating a model with two regimes and testing it against a model with three, at the significance level $\delta\varrho^2$. The sequence is terminated at the first non-rejection of the corresponding null hypothesis. The significance level is reduced at each step of the sequence and converges to zero, thereby avoiding excessively large models and controling the overall significance level. An upper bound for the overall significance level may be obtained using the Bonferroni bound (Gourieroux and Monfort 1995, p. 203). The selection of the parameter $\varrho$ is *ad hoc*. In order to avoid selecting small models (few regimes), it is good practice to carry the modeling cycle with different values of $\varrho$. In the empirical examples discussed in Section 8, we consider $\varrho = 1/2$ and $\varrho = 1/3$. The results are the same in both cases.

Evaluation following the estimation of the final model is performed by subjecting the model to the misspecification tests, as discussed in Lundbergh and Teräsvirta (2002).

---

[6]Bollerslev (1986) observed that under the null of homoskedasticity, there is no general Lagrange Multiplier test for GARCH($p$,$q$). This is due to the fact that the Hessian is singular if both $p > 0$ and $q > 0$.

## 7. MONTE-CARLO EXPERIMENT

The purpose of this section is to check the performance of the test described in Section 5. We use the following four data generating processes (DGPs):

(1) Model A:
   GARCH(1,1): $\alpha = 1.0 \times 10^{-5}$, $\beta = 0.85$, $\lambda = 0.05$.
(2) Model B:
   GARCH(1,1): $\alpha = 1.0 \times 10^{-5}$, $\beta = 0.90$, $\lambda = 0.088$.
(3) Model C:
   FCGARCH$(3, 1, 1)$: $\alpha_0 = 1 \times 10^{-4}$, $\beta_0 = 0.96$, $\lambda_0 = 0.18$, $\alpha_1 = -0.9 \times 10^{-4}$, $\beta_1 = -0.60$, $\lambda_1 = -0.10$, $\alpha_2 = 1 \times 10^{-4}$, $\beta_2 = 0.10$, $\lambda_2 = 0.05$, $\gamma_1 = 5000$, $\gamma_2 = 5000$, $c_1 = -0.005$, and $c_2 = 0.02$.
(4) Model D:
   FCGARCH$(3, 1, 1)$: $\alpha_0 = 6 \times 10^{-5}$, $\beta_0 = 1.10$, $\lambda_0 = 0.10$, $\alpha_1 = -5 \times 10^{-5}$, $\beta_1 = -0.65$, $\lambda_1 = -0.09$, $\alpha_2 = 1 \times 10^{-5}$, $\beta_2 = 0.10$, $\lambda_2 = 0.04$, $\gamma_1 = 3000$, $\gamma_2 = 3000$, $c_1 = -0.005$, and $c_2 = 0.005$.

In all DGPs the error term has a probability function either Gaussian or a standardized $t$ with 10 degrees of freedom. Model A has theoretical kurtosis 3.08 when the error distribution is Gaussian and 4.16 when the errors are $t$-distributed. Model B has a higher kurtosis: 8.55 with normality of the errors and 152.9 when the distribution of the errors is a $t$. Furthermore, model A has a well defined sixth-order moment even with $t$-distributed errors, while model B does not. We include model B in our simulation in order to evaluate the effect of the violation of the sixth-order moment assumption in the behavior of the test. Models C and D are different specifications of an FCGARCH$(3, 1, 1)$ model and were previously analyzed in the examples in Section 3. Using the result of Theorem 2, it can be shown that Models C and D satisfy the sixth-order moment condition. All the simulations are based on series with 1000 observations and the first 500 observations of each generated series are always discarded to avoid any initialization effect; see Lundbergh and Teräsvirta (2002). For each experiment, a total of 1000 replications have been generated. Only the results concerning the robust version of the tests are shown in order to save space.

Results from simulating the modeling strategy can be found in Table 2. The table also contains results on choosing the number of regimes using two information criteria: AIC and SBIC. The sequence of LM tests is carried out with three different initial significance levels $\alpha$. The value of the hyper-parameter $\varrho$ is $1/2$, meaning that at each step the significance level of the additional regime test is halved.

As can be seen from the table, both the AIC and the SBIC are very conservative, strongly underestimating the number of regimes in most of the cases. On the other hand, although still conservative, the sequence of LM tests selects the correct specification more often, specially in comparison with the former two information criteria. Another important fact is related to the

risk of specifying an overfitted model. It is clear from the table that, even with a large initial significance level (10%), overfitting occurs very rarely (less than 1% of the cases).

TABLE 2. SIMULATION: MODELING STRATEGY RESULTS.

The table reports the frequency that a model with a given number of limiting regimes is selected over 1000 simulations. 1000 observations of each model is simulated at each replication. In all the simulations the parameter $\varrho$ equals 2. $\delta$ is the initial significance level of sequence of LM tests.

| Model | Error Distribution | Number of Regimes | AIC | SBIC | LM test $(\delta = 0.01)$ | LM test $(\delta = 0.05)$ | LM test $(\delta = 0.10)$ |
|---|---|---|---|---|---|---|---|
| A | Gaussian | 1 | 1 | 1 | 1 | 0.960 | 0.904 |
| | | 2 | 0 | 0 | 0 | 0.036 | 0.088 |
| | | 3 | 0 | 0 | 0 | 0 | 0.004 |
| | | $\geq 4$ | 0 | 0 | 0 | 0.004 | 0.004 |
| | $t$ with 10 d.f. | 1 | 0.996 | 1 | 0.996 | 0.976 | 0.940 |
| | | 2 | 0.004 | 0 | 0.004 | 0.020 | 0.052 |
| | | 3 | 0 | 0 | 0 | 0 | 0.004 |
| | | $\geq 4$ | 0 | 0 | 0 | 0.004 | 0.004 |
| B | Gaussian | 1 | 1 | 1 | 0.992 | 0.956 | 0.908 |
| | | 2 | 0 | 0 | 0.008 | 0.040 | 0.080 |
| | | 3 | 0 | 0 | 0 | 0.004 | 0.004 |
| | | $\geq 4$ | 0 | 0 | 0 | 0 | 0.008 |
| | $t$ with 10 d.f. | 1 | 0.996 | 1 | 0.996 | 0.952 | 0.896 |
| | | 2 | 0.004 | 0 | 0.004 | 0.044 | 0.088 |
| | | 3 | 0 | 0 | 0 | 0.004 | 0.012 |
| | | $\geq 4$ | 0 | 0 | 0 | 0 | 0.004 |
| C | Gaussian | 1 | 0.016 | 0.016 | 0 | 0 | 0 |
| | | 2 | 0.952 | 0.984 | 0.956 | 0.904 | 0.828 |
| | | 3 | 0.032 | 0 | 0.036 | 0.092 | 0.164 |
| | | $\geq 4$ | 0 | 0 | 0.008 | 0.004 | 0.008 |
| | $t$ with 10 d.f. | 1 | 0.012 | 0.012 | 0.020 | 0.004 | 0 |
| | | 2 | 0.664 | 0.664 | 0.932 | 0.904 | 0.876 |
| | | 3 | 0.300 | 0.300 | 0.040 | 0.086 | 0.118 |
| | | $\geq 4$ | 0.024 | 0.024 | 0.008 | 0.006 | 0.006 |
| D | Gaussian | 1 | 0 | 0 | 0.012 | 0 | 0 |
| | | 2 | 0.964 | 0.976 | 0.880 | 0.760 | 0.652 |
| | | 3 | 0.032 | 0.024 | 0.098 | 0.236 | 0.346 |
| | | $\geq 4$ | 0.004 | 0 | 0.010 | 0.004 | 0.002 |
| | $t$ with 10 d.f. | 1 | 0.004 | 0.004 | 0.060 | 0.008 | 0.008 |
| | | 2 | 0.972 | 0.972 | 0.744 | 0.724 | 0.652 |
| | | 3 | 0.016 | 0.016 | 0.188 | 0.258 | 0.338 |
| | | $\geq 4$ | 0.008 | 0.008 | 0.008 | 0.010 | 0.002 |

## 8. EMPIRICAL EXAMPLES

We consider seven indexes: Amsterdam (EOE), Frankfurt (DAX), Hong Kong (Hang Seng), London (FTSE100), New York, (S&P 500), Paris (CAC40), and Tokyo (Nikkei). These indexes are chosen in order to represent some important financial markets. We split the sample into two parts. The first one is for in-sample analysis and the second one is used to test the forecasting performance of the models. For all series, except for the CAC40 index, the first sub-sample begins in January, 7 1986 and ends in December, 31 1997 (3128 observations). The CAC40 index begins in July, 9 1987 and ends in December, 31 1997, a total of 2736 observations. The second sub-sample begins in January, 5 1998 and ends in November, 11 2005 (2050 observations).

In order to correctly specify the conditional mean, we follow Engle and Ng (1993). The procedure involves a day-of-the-week effect adjustment and an autoregression which removes the linear predictable part of the daily returns. Let $y_t$ be the daily return at day $t$. We start regressing $y_t$ on a constant and four variables: $\text{Mon}_t$, $\text{Tue}_t$, $\text{Wed}_t$, and $\text{Thu}_t$, which are dummy variables for Monday, Tuesdays, Wednesdays, and Thursdays, respectively. The residual from the regression, $u_t$, is therefore regressed on a constant and on $u_{t-1}, \ldots, u_{t-7}$. We choose seven lags in order to remove any remaining day-of-the-week effect not captured by the dummy variables. The residual from the autoregression, $r_t$, is the unpredictable return. An alternative, frequently used in the literature, is to specify just a linear first-order linear autoregressive model for the returns. However, for the series considered in this paper this approach fails in removing the all the serial correlation in the returns, leading to a misspecified model for the conditional mean, which in turn, may lead to a misspecification of the conditional variance; see McAleer (2005, p. 247) for a nice discussion.

Table 3 shows the adjustment results. Table 4 shows descriptive statistics and diagnostics, where $\sigma$ is the standard deviation, $SK$ is the skewness, $K$ is the kurtosis, and $Q(10)$ and $QS(10)$ are, respectively, the $p$-values of the Ljung-Box statistic for tenth-order serial correlation in the unpredictable returns and squared returns. $Sb$, $Nsb$, $Psb$, and $Jsb$ are, respectively, the $p$-values of the sign bias, negative sign bias, positive sign bias, and joint tests for asymmetry proposed by Engle and Ng (1993). $ARCH(4)$ is the $p$-value of the fourth-order ARCH LM test described in Engle (1982). From the Ljung-Box test statistic at the 1% significance level we find no significant serial correlation left in the series after our adjustment procedure. The coefficients of skewness and kurtosis both indicate that the series have a distribution that is fat-tailed and skewed to the left. Furthermore, the Ljung-Box statistic in the squares and the ARCH LM test strongly suggest the presence of time-varying volatility. Moreover, there are evidence of asymmetries in the conditional variance of all the series. The negative sign bias and joint tests reject the null hypothesis of no asymmetric effect for all the eight indexes. The positive sign bias test strongly rejects the null hypothesis for the EOE, FTSE100, and CAC40 indexes. The sign bias test rejects the null for the DAX, Hang Seng, and Nikkei indexes. The overall evidence is that the size of negative past returns strongly affects the current volatility: Large negative unpredictable returns cause more volatility than small ones.

TABLE 3. DAILY RETURNS: MEAN ADJUSTMENT REGRESSIONS.

The table reports the results of a procedure to remove the day-of-the-week effects and the predictable part of the daily returns. Let $y_t$ be the daily return at time $t$. First, $y_t$ is regressed on a constant and four dummies representing the days-of-the week. The residual from this regression is regressed on a constant and seven lags to obtain the residual $r_t$, which is the unpredictable return. The figures between parentheses bellow the estimates are the Newey-West robust standard errors. The only exception is the CAC40 index, which begins in July, 9 1987 (2736 observations). The sample period is from January, 6 1986 to December, 31 1997 (3128 observations).

Day-of-the-week adjustment: $y_t = \theta_0 + \theta_1 Mon_t + \theta_2 Tue_t + \theta_3 Wed_t + \theta_4 Thu_t + u_t$

| Series | $\theta_0$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ |
|---|---|---|---|---|---|
| EOE | $9.91 \times 10^{-5}$ ($3.84\times10^{-4}$) | $-7.12 \times 10^{-4}$ ($7.27\times10^{-4}$) | $7.89 \times 10^{-4}$ ($5.85\times10^{-4}$) | $1.35 \times 10^{-3}$ ($6.06\times10^{-4}$) | $-7.34 \times 10^{-6}$ ($5.56\times10^{-4}$) |
| DAX | $4.81 \times 10^{-4}$ ($4.20\times10^{-4}$) | $-1.04 \times 10^{-4}$ ($7.37\times10^{-4}$) | $-3.12 \times 10^{-4}$ ($6.87\times10^{-4}$) | $7.72 \times 10^{-4}$ ($6.37\times10^{-4}$) | $-7.84 \times 10^{-5}$ ($6.12\times10^{-4}$) |
| Hang Seng | $1.42 \times 10^{-4}$ ($5.17\times10^{-4}$) | $-2.84 \times 10^{-3}$ ($1.07\times10^{-3}$) | $-3.31 \times 10^{-4}$ ($8.74\times10^{-4}$) | $7.54 \times 10^{-4}$ ($7.71\times10^{-4}$) | $-1.83 \times 10^{-3}$ ($8.23\times10^{-4}$) |
| FTSE100 | $7.88 \times 10^{-4}$ ($3.50\times10^{-4}$) | $-1.62 \times 10^{-4}$ ($5.06\times10^{-4}$) | $-2.37 \times 10^{-5}$ ($5.14\times10^{-4}$) | $1.62 \times 10^{-4}$ ($4.72\times10^{-4}$) | $-4.11 \times 10^{-4}$ ($4.63\times10^{-4}$) |
| S&P 500 | $2.74 \times 10^{-4}$ ($3.77\times10^{-4}$) | $1.45 \times 10^{-5}$ ($6.21\times10^{-4}$) | $7.80 \times 10^{-4}$ ($5.41\times10^{-4}$) | $6.00 \times 10^{-4}$ ($5.00\times10^{-4}$) | $-3.25 \times 10^{-4}$ ($4.96\times10^{-4}$) |
| CAC40 | $4.05 \times 10^{-4}$ ($4.51\times10^{-4}$) | $-2.14 \times 10^{-3}$ ($7.59\times10^{-4}$) | $4.78 \times 10^{-4}$ ($6.75\times10^{-4}$) | $4.64 \times 10^{-4}$ ($6.83\times10^{-4}$) | $4.62 \times 10^{-4}$ ($6.04\times10^{-4}$) |
| Nikkei | $-2.05 \times 10^{-4}$ ($4.96\times10^{-4}$) | $-1.39 \times 10^{-3}$ ($7.18\times10^{-4}$) | $5.26 \times 10^{-4}$ ($7.62\times10^{-4}$) | $9.04 \times 10^{-4}$ ($7.52\times10^{-4}$) | $1.24 \times 10^{-3}$ ($6.86\times10^{-4}$) |

Autocorrelation adjustment: $u_t = \phi_0 + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \phi_3 u_{t-3} + \phi_4 u_{t-3} + \phi_5 u_{t-5} + \phi_6 u_{t-6} + \phi_7 u_{t-7} + r_t$

| Series | $\phi_0$ | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ | $\phi_5$ | $\phi_6$ | $\phi_7$ |
|---|---|---|---|---|---|---|---|---|
| EOE | $3.81 \times 10^{-6}$ ($2.02\times10^{-4}$) | $-1.67 \times 10^{-2}$ ($4.56\times10^{-2}$) | $-3.56 \times 10^{-2}$ ($5.12\times10^{-2}$) | $3.34 \times 10^{-3}$ ($4.48\times10^{-2}$) | $6.88 \times 10^{-3}$ ($3.89\times10^{-2}$) | $3.05 \times 10^{-2}$ ($4.37\times10^{-2}$) | $-1.70 \times 10^{-2}$ ($4.04\times10^{-2}$) | $1.73 \times 10^{-2}$ ($3.78\times10^{-2}$) |
| DAX | $-1.69 \times 10^{-5}$ ($2.20\times10^{-4}$) | $-2.26 \times 10^{-3}$ ($3.39\times10^{-2}$) | $-3.94 \times 10^{-2}$ ($3.38\times10^{-2}$) | $5.98 \times 10^{-3}$ ($3.18\times10^{-2}$) | $-9.06 \times 10^{-3}$ ($2.86\times10^{-2}$) | $2.44 \times 10^{-2}$ ($2.96\times10^{-2}$) | $-3.95 \times 10^{-2}$ ($2.54\times10^{-2}$) | $2.51 \times 10^{-2}$ ($2.80\times10^{-2}$) |
| Hang Seng | $5.63 \times 10^{-6}$ ($3.00\times10^{-4}$) | $-2.32 \times 10^{-3}$ ($4.77\times10^{-2}$) | $-1.39 \times 10^{-2}$ ($3.02\times10^{-2}$) | $1.17 \times 10^{-1}$ ($3.34\times10^{-2}$) | $-1.57 \times 10^{-2}$ ($2.83\times10^{-2}$) | $3.82 \times 10^{-2}$ ($6.00\times10^{-2}$) | $-3.69 \times 10^{-2}$ ($2.64\times10^{-2}$) | $6.56 \times 10^{-3}$ ($3.34\times10^{-2}$) |
| FTSE100 | $7.60 \times 10^{-6}$ ($1.64\times10^{-4}$) | $7.42 \times 10^{-2}$ ($7.26\times10^{-2}$) | $-1.82 \times 10^{-3}$ ($5.31\times10^{-2}$) | $1.09 \times 10^{-2}$ ($3.84\times10^{-2}$) | $6.20 \times 10^{-2}$ ($3.61\times10^{-2}$) | $-1.46 \times 10^{-2}$ ($3.15\times10^{-2}$) | $-5.39 \times 10^{-3}$ ($2.78\times10^{-2}$) | $-1.48 \times 10^{-3}$ ($2.74\times10^{-2}$) |
| S&P 500 | $4.72 \times 10^{-6}$ ($1.77\times10^{-4}$) | $2.66 \times 10^{-2}$ ($6.18\times10^{-2}$) | $-6.03 \times 10^{-2}$ ($6.04\times10^{-2}$) | $-2.78 \times 10^{-2}$ ($5.38\times10^{-2}$) | $-4.16 \times 10^{-2}$ ($2.80\times10^{-2}$) | $4.15 \times 10^{-2}$ ($6.22\times10^{-2}$) | $-1.65 \times 10^{-2}$ ($3.11\times10^{-2}$) | $-8.74 \times 10^{-3}$ ($2.66\times10^{-2}$) |
| CAC40 | $6.17 \times 10^{-6}$ ($2.29\times10^{-4}$) | $2.20 \times 10^{-2}$ ($2.92\times10^{-2}$) | $2.43 \times 10^{-2}$ ($3.90\times10^{-2}$) | $-2.15 \times 10^{-2}$ ($3.00\times10^{-2}$) | $4.12 \times 10^{-3}$ ($2.79\times10^{-2}$) | $-1.20 \times 10^{-2}$ ($8.30\times10^{-2}$) | $3.39 \times 10^{-3}$ ($2.48\times10^{-2}$) | $1.67 \times 10^{-2}$ ($3.32\times10^{-2}$) |
| Nikkei | $3.52 \times 10^{-6}$ ($2.42\times10^{-4}$) | $-3.60 \times 10^{-3}$ ($3.57\times10^{-2}$) | $-7.36 \times 10^{-2}$ ($2.61\times10^{-2}$) | $7.70 \times 10^{-3}$ ($2.74\times10^{-2}$) | $3.01 \times 10^{-2}$ ($2.81\times10^{-2}$) | $2.48 \times 10^{-2}$ ($2.47\times10^{-2}$) | $-1.99 \times 10^{-2}$ ($2.32\times10^{-2}$) | $-9.14 \times 10^{-3}$ ($2.45\times10^{-2}$) |

TABLE 4. DAILY UNPREDICTABLE RETURNS: DESCRIPTIVE STATISTICS AND DIAGNOSTICS.

The table shows descriptive statistics and diagnostics for the unpredictable daily returns. $\sigma$ is the standard deviations, $SK$ is the skewness, $K$ is the kurtosis, $Q(10)$ is the $p$-value of the Ljung-Box statistic for tenth-order serial correlation in the unpredictable returns, $QS(10)$ is the $p$-value of the Ljung-Box statistic for tenth-order serial correlation in the unpredictable squared returns, and $Sb$, $Nsb$, $Psb$, and $Jsb$ are, respectively, the $p$-values of the sign bias, negative sign bias, positive sign bias, and joint tests for asymmetry proposed by Engle and Ng (1993). $ARCH(4)$ is the $p$-value of the fourth-order ARCH LM test described in Engle (1982).

| Series | $\sigma$ | $SK$ | $K$ | $Q(10)$ | $QS(10)$ | $Sb$ | $Nsb$ | $Psb$ | $Jsb$ | $ARCH(4)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| EOE | 0.011 | -0.75 | 19.31 | 0.02 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 |
| DAX | 0.012 | -0.99 | 14.93 | 0.99 | 0.00 | 0.01 | 0.00 | 0.12 | 0.00 | 0.00 |
| Hang Seng | 0.016 | -4.84 | 115.71 | 0.43 | 0.00 | 0.03 | 0.00 | 0.65 | 0.00 | 0.02 |
| FTSE100 | 0.009 | -1.34 | 25.23 | 0.45 | 0.00 | 0.74 | 0.00 | 0.00 | 0.00 | 0.00 |
| S&P 500 | 0.010 | -4.44 | 100.71 | 0.99 | 0.00 | 0.16 | 0.00 | 0.43 | 0.00 | 0.00 |
| CAC40 | 0.012 | -0.46 | 10.17 | 0.62 | 0.00 | 0.52 | 0.00 | 0.01 | 0.00 | 0.00 |
| Nikkei | 0.014 | -0.25 | 14.77 | 0.18 | 0.00 | 0.01 | 0.00 | 0.42 | 0.00 | 0.00 |

Using the unpredictable return series, we estimate the standard GARCH(1,1) model, as well as the GJR-GARCH(1,1) specification. The estimation is performed using the Bollerslev-Wooldridge quasi-maximum likelihood approach and the Marquardt algorithm. The adequacy of these models is then checked using the sign bias, negative sign bias and positive sign bias tests. Table 5 reports the estimation and diagnostic test results of GARCH(1,1) and GJR-GARCH(1,1) models for the daily unpredictable returns. The number in parentheses below the estimates are the Bollerslev-Wooldridge robust standard errors. $Pc$, $Pi$, and $Pcc$ are the $p$-values of the tests of unconditional coverage, independence, and conditional coverage proposed by Christoffersen (1998) to evaluate interval estimation. In the present case a 95% confidence interval is considered.

By inspection of Table 5 it is clear that, with the exception of the CAC40 index, the normalized residuals from GJR-GARCH(1,1) have lower kurtosis than the ones from the GARCH(1,1) alternative. The skewness coefficients are also lower for the GJR-GARCH(1,1) model. Several other interesting facts emerge from the table. First, the sum of the estimated $\beta_0$ and $\lambda_0$ coefficients in the GARCH(1,1) models is over 0.94 for all series, indicating a high persistence in the dynamics of the estimated volatility. For all the series, the coefficients are statistically significant at the 5% level. Concerning the results of the sign-bias, negative sing-bias, positive sign-bias, and joint tests it is clear from the analysis of the results in Table 5 that there are still asymmetric effects in the normalized residuals from the GARCH(1,1) models. The only case where the test statistics is not significant are the FTSE100. The analysis of the coverage tests indicates that the GARCH(1,1) fails to produce correct confidence intervals for three of the series considered: DAX, Hang Seng, and FTSE100.

When the GJR-GARCH(1,1) model is considered, it is important to mention that a negative shock induces an explosive regime, as the sum of the estimated $\beta_0$, $\lambda_0$, and $\lambda_1$ parameters is greater than one for all series, with the only exception of the CAC40 index; See Table 7. The parameter $\lambda_1$ is significant for all series except from the FTSE100 and S&P500. Concerning the results of

the sign bias, negative sign bias, positive sign bias, and joint tests it seems that there are still some asymmetric effects in three out of the eight series considered here, namely: S&P500, CAC40, and Nikkei indexes. Finally, the results of the coverage tests indicate that the GJR-GARCH(1,1) model does not provide correct interval estimation for the DAX, and FTSE100 series. One may argue that if a $t$-distribution is considered instead of the Gaussian one, the coverage probability of the GARCH and GJR-GARCH may be improved. However, as pointed out in Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2001a, 2001b, 2003) the distribution of the standardized returns are nearly Gaussian. For that reason, we decided to keep the normality assumption in order to check if the presence of more than two regimes in the dynamics of the conditional variance is one of the causes of the remaining excess of kurtosis and poor coverage probabilities.

We proceed specifying an FCGARCH model having the GARCH(1,1) specification as our basis model. Applying the robust version of the LM test developed in Section 5 the null hypothesis is rejected for all series with the only exception of the FTSE100 index. At each step of the testing sequence we halve the significance level of the test ($\varrho = 1/2$). We also carry the test sequence with other values for $\varrho$ and the results do not change. The initial significance level for the sequence of LM tests is 5%. Table 6 shows the estimation results and diagnostic statistics. The estimation is performed by the quasi maximum likelihood method using the Sequential Quadratic Programming numerical optimization algorithm. To avoid convergence problems, we divide the transition variable, $r_{t-1}$, by its unconditional standard deviation. The number in parentheses below the estimates are the standard errors.

The sequence of robust LM tests shows evidence of two limiting regimes for three series: EOE, Hang Seng, and Nikkei indexes. It is important to mention that for the EOE and Nikkei series the parameter $c_1$ is positive and statistically different from zero which contradicts the usual zero threshold considered in the literature. For the Hang Seng the result is opposed: the parameter $c_1$ is not statistically different from zero, corroborating previous results. It is important to mention that comparing the AIC from the FCGARCH model with the one from the GARCH and GJR-GARCH specifications, the FCGARCH outperforms the other two alternatives, indicating that the final model is not overparametrized.

For the DAX, S&P500, and CAC40 three limiting regimes are found. It is clear that for all the three series the first limiting (extreme) regime is associated with very negative shocks, representing great losses. The middle regime is related to tranquil periods and the third and extreme regime represents large positive shocks.

Observing the results in Table 6 it is clear that the estimated standardized residuals from FCGARCH model have kurtosis coefficients lower than both the GARCH(1,1) and GJR-GARCH(1,1) models. For example, for the DAX index, the reduction in the estimated kurtosis is about 50% when compared to the GJR-GARCH alternative. In addition, the standardized residuals from the FCGARCH model are less skewed than the ones from the GARCH and GJR-GARCH models.

The only exception is the Nikkei index, for which the GJR-GARCH specification has least skewed normalized residuals. As in the GJR-GARCH(1,1) case, the FCGARCH model seems to describe adequately the asymmetric relation between returns and volatility, with the exception of the S&P500, CAC40, and Nikkei series. For those series, a higher order model may be more adequate. However, this investigation is beyond the scope of the paper. We do not report the standard errors for the slope parameters because they are not very accurate as the magnitude of the estimated $\gamma$s are very high, indicating a sharp transition among regimes. Moreover, as pointed out in Section 5, the $t$-statistic does not have its customary distribution under the null hypothesis that $\gamma = 0$. In addition, when the coverage tests are considered, the FCGARCH model seems to outperform the concurrent models considered in the paper and produces correct confidence intervals for all the series.

One very interesting fact is the large value of the estimated $\beta_0$s, which indicates a very persistent regime associated with negative returns. Table 7 shows the persistence associated with each limiting regime in both the GJR-GARCH and FCGARCH models. Considering the GJR-GARCH model, the sum $\beta_0 + \lambda_0 + \lambda_1$ is the persistence associated with negative past returns ("bad news"), whereas the $\beta_0 + \lambda_0$ represents the persistence when the past return is positive ("good news"). On the other hand, in the FCGARCH specification, the sum $\beta_0 + \lambda_0$ is the persistence in the first extreme regime that can be associated with "bad" or "very bad" news depending if the estimated model has two or three limiting regimes. The sum $\beta_0 + \beta_1 + \lambda_0 + \lambda_1$ is the persistence either in the "tranquil period" or in "very good news regime". Finally the last column in the table shows the persistence of last limiting regime in the FCGARCH model and is associated with "good" or "very good" news depending if the estimated model has two or three regimes. Some interesting facts emerge from the table. First, the regime associated with negative returns is much more persistent in the FCGARCH model than in the GJR-GARCH specification. Second, the GARCH effect seems to be dissipated when the returns become more positive, specially when there are three regimes and not only two. Finally, even with a very high persistent regime, all the models are stationary, as restriction (7) is met for all cases.

Finally, we test the forecasting performance of the estimated FCGARCH models. We use the mean absolute errors as a performance measure. The squared returns are used as a proxy to the volatility. The results are shown in Table 8. Analyzing the results, we can observe that apart from the S&P 500 case, the FCGARCH model performs slightly better than the other two specifications.

TABLE 5. ESTIMATED MODELS AND SPECIFICATION TESTS.

The table reports the estimation and diagnostic test results of GARCH(1,1) and GJR-GARCH(1,1) models for the unpredictable daily returns. The estimation is performed by the quasi maximum likelihood method. The number in parentheses below the estimates are the Bollerslev-Wooldridge robust standard errors. $\sigma$, $SK$ and $K$ are the standard deviation, the skewness and the kurtosis of the standardized residuals, respectively. $Sb$, $Nsb$, $Psb$, and $Jsb$ are the p-values of the sign bias, negative sign bias, positive sign bias, and joint tests for asymmetry proposed in Engle and Ng (1993). $Pc$, $Pi$, and $Pcc$ are the p-values of the tests of unconditional coverage, independence, and conditional 95% coverage. AIC is the value of the Akaike's information criterium and Likelihood is the estimated log-likelihood.

GARCH(1,1) model: $h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 r_{t-1}^2$

| Series | $\alpha_0$ | $\beta_0$ | $\lambda_0$ | $\sigma$ | $K$ | $SK$ | $Sb$ | $Nsb$ | $Psb$ | $Jsb$ | $Pc$ | $Pi$ | $Pcc$ | AIC | Likelihood |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EOE | $2.32 \times 10^{-6}$ $(8.15 \times 10^{-7})$ | 0.91 (0.02) | 0.07 (0.03) | 1.00 | 12.61 | -1.16 | 0.02 | 0.48 | 0.02 | 0.08 | 0.93 | 0.50 | 0.80 | -4.16 | $1.01 \times 10^4$ |
| DAX | $7.85 \times 10^{-6}$ $(3.29 \times 10^{-6})$ | 0.82 (0.04) | 0.14 (0.05) | 1.00 | 17.39 | -1.34 | 0.04 | 0.35 | 0.03 | 0.14 | 0.02 | 0.88 | 0.08 | -4.01 | $9.64 \times 10^3$ |
| Hang Seng | $1.14 \times 10^{-5}$ $(5.00 \times 10^{-6})$ | 0.79 (0.03) | 0.17 (0.05) | 1.00 | 18.58 | -1.55 | 0.01 | 0.08 | 0.05 | 0.06 | 0.68 | 0.02 | 0.06 | -3.82 | $9.06 \times 10^3$ |
| FTSE100 | $4.86 \times 10^{-6}$ $(2.29 \times 10^{-6})$ | 0.84 (0.05) | 0.10 (0.03) | 0.99 | 17.22 | -1.06 | 0.48 | 0.98 | 0.29 | 0.16 | 0.00 | 0.92 | 0.00 | -4.28 | $1.05 \times 10^4$ |
| S&P 500 | $1.73 \times 10^{-6}$ $(5.94 \times 10^{-7})$ | 0.89 (0.04) | 0.09 (0.05) | 1.00 | 11.81 | -1.06 | 0.01 | 0.00 | 0.01 | 0.01 | 0.62 | 0.88 | 0.88 | -4.29 | $1.05 \times 10^4$ |
| CAC40 | $7.97 \times 10^{-6}$ $(2.72 \times 10^{-6})$ | 0.83 (0.03) | 0.11 (0.02) | 1.00 | 6.30 | -0.46 | 0.08 | 0.79 | 0.04 | 0.06 | 0.51 | 0.96 | 0.81 | -4.02 | $8.45 \times 10^3$ |
| Nikkei | $3.91 \times 10^{-6}$ $(1.42 \times 10^{-6})$ | 0.84 (0.04) | 0.16 (0.06) | 1.00 | 15.07 | -0.76 | 0.05 | 0.00 | 0.02 | 0.00 | 0.56 | 0.73 | 0.80 | -3.95 | $9.46 \times 10^3$ |

GJR-GARCH(1,1) model: $h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 r_{t-1}^2 + \lambda_1 r_{t-1}^2 I(r_{t-1} \leq 0)$

| Series | $\alpha_0$ | $\beta_0$ | $\lambda_0$ | $\lambda_1$ | $\sigma$ | $K$ | $SK$ | $Sb$ | $Nsb$ | $Psb$ | $Jsb$ | $Pc$ | $Pi$ | $Pcc$ | AIC | Likelihood |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EOE | $2.63 \times 10^{-6}$ $(7.09 \times 10^{-7})$ | 0.91 (0.02) | 0.03 (0.02) | 0.07 (0.03) | 1.00 | 10.62 | -0.98 | 0.07 | 0.87 | 0.10 | 0.10 | 0.81 | 0.42 | 0.70 | -4.16 | $1.01 \times 10^4$ |
| DAX | $9.11 \times 10^{-6}$ $(3.19 \times 10^{-6})$ | 0.82 (0.04) | 0.04 (0.02) | 0.15 (0.07) | 1.00 | 15.64 | -1.17 | 0.23 | 0.96 | 0.30 | 0.49 | 0.03 | 0.82 | 0.11 | -4.02 | $9.67 \times 10^3$ |
| Hang Seng | $1.41 \times 10^{-5}$ $(3.64 \times 10^{-6})$ | 0.79 (0.02) | 0.04 (0.03) | 0.24 (0.09) | 1.00 | 12.48 | -1.01 | 0.08 | 0.90 | 0.78 | 0.18 | 0.93 | 0.97 | 1.00 | -3.84 | $9.11 \times 10^3$ |
| FTSE100 | $5.06 \times 10^{-6}$ $(1.85 \times 10^{-6})$ | 0.83 (0.04) | 0.04 (0.02) | 0.12 (0.08) | 0.99 | 12.92 | -0.78 | 0.23 | 0.32 | 0.85 | 0.32 | 0.00 | 0.06 | 0.00 | -4.29 | $1.05 \times 10^4$ |
| S&P 500 | $2.52 \times 10^{-6}$ $(8.33 \times 10^{-7})$ | 0.88 (0.03) | 0.04 (0.02) | 0.09 (0.05) | 1.00 | 10.37 | -0.94 | 0.03 | 0.13 | 0.03 | 0.10 | 0.42 | 0.76 | 0.69 | -4.30 | $1.05 \times 10^4$ |
| CAC40 | $8.69 \times 10^{-6}$ $(2.70 \times 10^{-6})$ | 0.85 (0.03) | 0.03 (0.01) | 0.12 (0.03) | 1.00 | 6.49 | -0.44 | 0.33 | 0.16 | 0.48 | 0.07 | 0.51 | 0.71 | 0.75 | -4.02 | $8.47 \times 10^3$ |
| Nikkei | $3.87 \times 10^{-6}$ $(1.18 \times 10^{-6})$ | 0.85 (0.02) | 0.04 (0.02) | 0.21 (0.08) | 1.00 | 11.16 | -0.45 | 0.50 | 0.03 | 0.22 | 0.03 | 0.68 | 0.90 | 0.91 | -3.96 | $9.51 \times 10^3$ |

TABLE 6. FCGARCH ESTIMATION AND SPECIFICATION TESTS.

The table reports the estimation and diagnostic test results of the FCGARCH(1,1) model for the unpredictable daily returns. The estimation is performed by the quasi maximum likelihood method. The number in parentheses below the estimates are the standard errors. $\sigma$, $SK$ and $K$ are the standard deviation, the skewness and the kurtosis of the standardized residuals, respectively. $Sb$, $Nsb$, $Psb$, and $Jsb$ are the $p$-values of the Sign bias, negative sign bias, positive sign bias, and joint tests for asymmetry proposed in Engle and Ng (1993). $Pc$, $Pi$, and $Pcc$ are the $p$-values of the tests of unconditional coverage, independence, and conditional coverage. In the present case a 95% confidence interval is considered. AIC is the value of the Akaike's information criterium and Likelihood is the estimated log-likelihood. The estimated model has the following form:
$$h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 g_{t-1}^2 + \left(\alpha_1 + \beta_1 h_{t-1} + \lambda_1 r_{t-1}^2\right) f\left(r_{t-1}/\sigma; \gamma_1, c_1\right) + \left(\alpha_2 + \beta_2 h_{t-1} + \lambda_2 r_{t-1}^2\right) f\left(r_{t-1}/\sigma; \gamma_2, c_2\right),$$
where $\sigma$ is the unconditional standard deviation of $r_{t-1}$.

### Estimation results

| Series | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\lambda_0$ | $\lambda_1$ | $\lambda_2$ | $\gamma_1$ | $\gamma_2$ | $c_1$ | $c_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EOE | $1.010 \times 10^{-8}$ <br>$(2.575\times10^{-5})$ | $1.085 \times 10^{-5}$ <br>$(4.414\times10^{-5})$ | — | 0.97 <br>(0.08) | -0.28 <br>(0.10) | — | 0.13 <br>(0.05) | -0.07 <br>(0.07) | — | 78.88 <br>(—) | — | 0.22 <br>(0.07) | — |
| DAX | $4.70 \times 10^{-6}$ <br>$(1.47\times10^{-4})$ | $1.50 \times 10^{-7}$ <br>$(4.91\times10^{-5})$ | $6.20 \times 10^{-5}$ <br>$(5.06\times10^{-4})$ | 1.13 <br>(0.19) | -0.26 <br>(0.21) | -0.41 <br>(0.20) | 0.17 <br>(0.13) | -0.17 <br>(0.22) | 0 <br>(0.24) | 73.70 <br>(—) | 73.69 <br>(—) | -0.99 <br>(0.05) | 0.76 <br>(0.11) |
| Hang Seng | $2.22 \times 10^{-16}$ <br>$(1.90\times10^{-3})$ | $3.48 \times 10^{-5}$ <br>$(3.60\times10^{-3})$ | — | 0.96 <br>(0.08) | -0.40 <br>(0.08) | — | 0.29 <br>(0.05) | -0.22 <br>(0.08) | — | 46.94 <br>(—) | — | 0.03 <br>(0.04) | — |
| FTSE100 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| S&P 500 | $2.13 \times 10^{-16}$ <br>$(2.51\times10^{-3})$ | $2.123 \times 10^{-7}$ <br>$(2.500\times10^{-3})$ | $-2.123 \times 10^{-7}$ <br>$(1.110\times10^{-3})$ | 1.67 <br>(0.17) | -0.71 <br>(0.17) | -0.82 <br>(2.00) | 0.01 <br>(0.00) | 0.03 <br>(0.01) | 0.13 <br>(0.06) | 39.48 <br>(—) | 39.48 <br>(—) | -2.14 <br>(0.00) | 1.83 <br>(0.01) |
| CAC40 | $2.23 \times 10^{-16}$ <br>$(5.01\times10^{-4})$ | $3.60 \times 10^{-6}$ <br>$(5.02\times10^{-4})$ | $6.48 \times 10^{-5}$ <br>$(3.20\times10^{-4})$ | 1.14 <br>(0.09) | -0.24 <br>(0.09) | -0.43 <br>(0.06) | 0.12 <br>(0.08) | -0.08 <br>(0.10) | -0.04 <br>(0.01) | 71.68 <br>(—) | 71.68 <br>(—) | -1.35 <br>(0.01) | 1.15 <br>(0.05) |
| Nikkei | $2.23 \times 10^{-16}$ <br>$(2.25\times10^{-5})$ | $5.69 \times 10^{-6}$ <br>$(3.12\times10^{-5})$ | — | 0.93 <br>(0.09) | -0.24 <br>(0.08) | — | 0.24 <br>(0.10) | -0.15 <br>(0.08) | — | 67.48 <br>(—) | — | 0.44 <br>(0.09) | — |

### Diagnostic test results

| Series | $\sigma$ | $K$ | $SK$ | $Sb$ | $Nsb$ | $Psb$ | $Jsb$ | $Pc$ | $Pi$ | $Pcc$ | $AIC$ | Likelihood |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EOE | 1.00 | 8.74 | -0.81 | 0.44 | 0.21 | 0.18 | 0.11 | 0.37 | 0.27 | 0.37 | -4.17 | $1.02 \times 10^4$ |
| DAX | 1.00 | 10.83 | -0.87 | 0.30 | 0.64 | 0.25 | 0.35 | 0.18 | 0.49 | 0.32 | -4.03 | $9.72 \times 10^3$ |
| Hang Seng | 1.00 | 10.51 | -0.79 | 0.51 | 0.66 | 0.80 | 0.55 | 0.87 | 0.60 | 0.86 | -3.85 | $9.14 \times 10^3$ |
| FTSE 100 | — | — | — | — | — | — | — | — | — | — | — | — |
| S&P500 | 1.00 | 8.80 | -0.81 | 0.01 | 0.04 | 0.04 | 0.03 | 0.20 | 0.61 | 0.38 | -4.30 | $1.06 \times 10^4$ |
| CAC40 | 1.00 | 5.55 | -0.39 | 0.21 | 0.19 | 0.30 | 0.04 | 0.84 | 0.57 | 0.84 | -4.02 | $8.48 \times 10^3$ |
| Nikkei | 1.00 | 10.86 | -0.51 | 0.66 | 0.02 | 0.16 | 0.03 | 0.80 | 0.14 | 0.33 | -3.97 | $9.53 \times 10^3$ |

TABLE 7. GJR-GARCH AND FCGARCH MODELS: PERSISTENCE IN EACH REGIME.

The table shows the persistence associated with each limiting regime in both the GJR-GARCH and FCGARCH models. The sum $\beta_0 + \lambda_0 + \lambda_1$ is the persistence associated with negative past returns in the GJR-GARCH model ("bad news"), whereas the $\beta_0 + \lambda_0$ represents the persistence when the past return is positive ("good news"). On the other hand, in the FCGARCH model, the sum $\beta_0 + \lambda_0$ is the persistence in the first extreme regime that can be associated with "bad" or "very bad" news depending if the estimated model has two or three limiting regimes. The sum $\beta_0 + \beta_1 + \lambda_0 + \lambda_1$ is the persistence either in the "tranquil period" or in the "very good news regime". Finally, the last column in the table shows the persistence of last limiting regime in the FCGARCH model and is associated with "good" or "very good" news depending if the estimated model has two or three regimes.

| | GJR-GARCH(1,1) model | | FCGARCH(1,1) model | | |
| | "Bad news" | "Good news" | $\beta_0 + \lambda_0$ | $\beta_0 + \beta_1$ | $\beta_0 + \beta_1 + \beta_2$ |
| Series | $\beta_0 + \lambda_0 + \lambda_1$ | $\beta_0 + \lambda_0$ | | $+\lambda_0 + \lambda_1$ | $+\lambda_0 + \lambda_1 + \lambda_2$ |
|---|---|---|---|---|---|
| EOE | 1.01 | 0.76 | 1.10 | 0.75 | – |
| DAX | 1.02 | 0.86 | 1.30 | 0.87 | 0.46 |
| Hang Seng | 1.07 | 0.83 | 1.25 | 0.62 | – |
| S&P 500 | 1.02 | 0.93 | 1.68 | 1.00 | 0.31 |
| CAC40 | 0.99 | 0.87 | 1.27 | 0.95 | 0.48 |
| Nikkei | 1.10 | 0.90 | 1.17 | 0.77 | – |

TABLE 8. FORECASTING PERFORMANCE: MEAN ABSOLUTE ERRORS.

The table shows the mean absolute errors for the one-step-ahead forecasts computed with different models. All the figures should be multiplied by $10^{-4}$. The "actual" volatility proxy is the squared returns. The forecasting period is from January, 5 1998 to November, 11 2005 (2050 observations).

| Series | GARCH(1,1) | GJR-GARCH(1,1) | FC-GARCH($m$,1,1) |
|---|---|---|---|
| EOE | 2.35 | 2.30 | 2.25 |
| DAX | 2.70 | 2.63 | 2.55 |
| Hang Seng | 2.99 | 2.89 | 2.88 |
| S&P 500 | 1.45 | 1.40 | 1.41 |
| CAC40 | 2.16 | 2.11 | 2.08 |
| Nikkei | 2.32 | 2.32 | 2.31 |

## 9. CONCLUSIONS

In this paper we put forward a new nonlinear GARCH(1,1) model to describe the asymmetric behavior observed in financial time series, as well as intermittent dynamics and excess of kurtosis. The model is called the Flexible Coefficient Smooth Transition GARCH (FCGARCH) and is a straightforward generalization of the Logistic Smooth Transition GARCH (LST-GARCH) model, being capable of modeling multiple regimes in the conditional variance of the series. The proposed model describes some of the stylized facts of financial time-series that existing techniques fail to model satisfactorily. Conditions for strict stationarity and ergodicity of the proposed model was established and the existence of the second- and fourth-order moments was carefully discussed.

It was shown that the model may have explosive regimes and still be strictly stationary and ergodic. Furthermore, estimation of the parameters was addressed and the asymptotic properties of the quasi-maximum likelihood estimator was derived under second- and fourth-order moment conditions. A modeling cycle based on a sequence of simple and easily implemented Lagrange multiplier tests is discussed in order to avoid the estimation of unidentified models. A Monte-Carlo experiment is designed to evaluate the methodology and it was shown that the modeling strategy works well in moderate samples.

An empirical example with seven stock indexes showed that the FCGARCH model was able to produce normalized residuals with lower kurtosis than the GARCH and GJR-GARCH models. Moreover, the results showed evidence of two limiting regimes for three series and three limiting regimes for other three. Only for one stock index there was no evidence of more than one regime. In addition, for all the series with three limiting regimes, the first limiting (extreme) regime was associated with very negative shocks, representing great losses. The middle regime was related to tranquil periods and the third and extreme regime represented large positive shocks. Thus we found strong evidence of both size and sign asymmetries. The first limiting regime for seven of the series was extremely explosive indicating that bad news may induce very high volatility. When a forecasting exercise was considered, the FCGARCH slightly outperformed the GARCH and GJR-GARCH alternatives.

## 10. ACKNOWLEDGMENTS

## APPENDIX A. PROOFS OF THEOREMS AND COROLLARIES

A.1. **Proof of Theorem 1.** The conditional variance $h_t$ in (1) can be written as

$$(A.1) \qquad h_t = g_{t-1} + \sum_{k=1}^{t-1} \prod_{j=0}^{k-1} c_{t-1-j} g_{t-k-1} + \prod_{j=0}^{t-1} c_{t-1-j} h_0.$$

Under Assumptions 4 and 5, $h_0 > 0$ with probability one. Furthermore, it is clear that there is a positive and finite constant $M$, such that $g_t \geq M$ with probability one. Then,

$$(A.2) \qquad h_t \geq M \left[ \sup_{1 \leq k \leq t-1} \prod_{j=0}^{k-1} c_{t-1-j} \right].$$

As the functions $f_{i,t}$, $i = 1, \ldots, H$, are bounded and $\varepsilon_t \sim \mathsf{IID}(0,1)$, it is easy to show that the sequence $\{c_t\}$ is strongly stationary and ergodic with $\mathsf{E}\left[|c_t|^{1+\delta}\right] < \infty$, $\forall t$ and for any $\delta$ arbitrarily close to zero. In addition, following the same arguments as in Corollary 1 in Trapletti, Leisch, and Hornik (2000), it is straightforward to show that $\{c_t\}$ is also $\alpha$-mixing with size $-a$, for any $a \in \mathbb{R}$, such that the law of large numbers for dependent and heterogeneously distributed observations applies (White 2001, Corollary 3.48, p.49). Hence, the remaining of the proof is identical of the one of Theorem 2 in Nelson (1990). This completes the proof.

*Q.E.D*

A.2. **Proof of Corollary 1.** Consider a positive constant $N < \infty$ and an indicator function defined as

$$(A.3) \qquad I_{h_{t-1} \geq N} = \begin{cases} 1 & \text{if } h_{t-1} \geq N \\ 0 & \text{otherwise.} \end{cases}$$

Set $I_{h_{t-1} < N} = 1 - I_{h_{t-1} \geq N}$.

Note that, since

$$\lim_{y_{t-1} \to \infty} \left[ \sum_{i=1}^{H} \lambda_i f_{i,t-1} \right] = \sum_{i=1}^{H} \lambda_i \equiv \lambda_U, \quad \lim_{y_{t-1} \to -\infty} \left[ \sum_{i=1}^{H} \lambda_i f_{i,t-1} \right] = 0,$$

$$\lim_{y_{t-1} \to \infty} \left[ \sum_{i=1}^{H} \beta_i f_{i,t-1} \right] = \sum_{i=1}^{H} \beta_i \equiv \beta_U \quad \text{and} \quad \lim_{y_{t-1} \to -\infty} \left[ \sum_{i=1}^{H} \beta_i f_{i,t-1} \right] = 0,$$

there will always exist a finite constant $M > 0$ and small numbers $\delta_\lambda > 0$ and $\delta_\beta > 0$ such that $\left| \left( \sum_{i=1}^{H} \lambda_i f_{i,t-1} \right) - \lambda_U \right| \leq \delta_\lambda$ and $\left| \left( \sum_{i=1}^{H} \beta_i f_{i,t-1} \right) - \beta_U \right| \leq \delta_\beta$, if $y_{t-1} \geq M$ and $\left| \sum_{i=1}^{H} \lambda_i f_{i,t-1} \right| \leq \delta_\lambda$ and $\left| \sum_{i=1}^{H} \beta_i f_{i,t-1} \right| \leq \delta_\beta$, if $y_{t-1} < -M$.

Take a large value for the constants $M$ and $N$ and write the following expected value.

$$\mathsf{E}\left[\log(c_{t-1})\right] = \mathsf{E}\left[ \log(c_{t-1}) \middle| I_{|\varepsilon_{t-1}| < \frac{M}{h_{t-1}^{-1/2}}} \right] \mathsf{Pr}\left[ I_{|\varepsilon_{t-1}| < \frac{M}{h_{t-1}^{-1/2}}} \right]$$

$$+ \mathsf{E}\left[ \log(c_{t-1}) \middle| I_{|\varepsilon_{t-1}| \geq \frac{M}{h_{t-1}^{-1/2}}} \right] \mathsf{Pr}\left[ I_{|\varepsilon_{t-1}| \geq \frac{M}{h_{t-1}^{-1/2}}} \right]$$

$$= \mathsf{E}\left[ \log(c_{t-1}) \middle| I_{|\varepsilon_{t-1}| < \frac{M}{h_{t-1}^{-1/2}}} \right] p + \mathsf{E}\left[ \log(c_{t-1}) \middle| I_{|\varepsilon_{t-1}| \geq \frac{M}{h_{t-1}^{-1/2}}} \right] (1-p).$$

An upper bound for $p$ can be made arbitrarily small by increasing $N$. As a consequence, for any small number $\delta > 0$ and large $M$, there is a constant $N$ such that

$$\mathsf{E}\left[\log(c_{t-1})\right] \leq \delta + \mathsf{E}\left[ \log(c_{t-1}) \middle| I_{|\varepsilon_{t-1}| \geq \frac{M}{h_{t-1}^{-1/2}}} \right] (1-p).$$

The log-moment condition can be expressed as

$$\mathsf{E}\left[\log(c_{t-1})\bigg|\, I_{|\varepsilon_{t-1}|\geq \frac{M}{h_{t-1}^{-1/2}}}\right] < 0.$$

However,

$$\mathsf{E}\left[\log(c_{t-1})\bigg|\, I_{|\varepsilon_{t-1}|\geq \frac{M}{h_{t-1}^{-1/2}}}\right] < \log\left\{\mathsf{E}\left[(c_{t-1})\bigg|\, I_{|\varepsilon_{t-1}|\geq \frac{M}{h_{t-1}^{-1/2}}}\right]\right\}.$$

Then, to satisfy the log-moment condition is sufficient that

$$\mathsf{E}\left[(c_{t-1})\bigg|\, I_{|\varepsilon_{t-1}|\geq \frac{M}{h_{t-1}^{-1/2}}}\right] \leq 1.$$

Then, as $\delta_\lambda$ and $\delta_\beta$ can be made arbitrarily small,

$$\mathsf{E}\left[(c_{t-1})\bigg|\, I_{|\varepsilon_{t-1}|\geq \frac{M}{h}}\right] = \frac{\mathsf{E}\left[\beta_U + \varepsilon_{t-1}^2\lambda_U | I_{|\varepsilon_{t-1}|\geq \frac{M}{h_{t-1}^{-1/2}}}\right] + \mathsf{E}\left[\beta_0 + \varepsilon_{t-1}^2\lambda_0\bigg|\, I_{|\varepsilon_{t-1}|< \frac{M}{h_{t-1}^{-1/2}}}\right]}{2}$$

$$\leq \frac{\beta_U + \lambda_U + \beta_0 + \lambda_0}{2}$$

Then, if

$$\frac{1}{2}\left(\beta_0 + \lambda_0\right) + \frac{1}{2}\sum_{i=0}^{H}\left(\beta_i + \lambda_i\right) \leq 1$$

is sufficient to guarantee the log-moment condition.

*Q.E.D*

A.3. **Proof of Theorem 2.** First set $y_t^{2k} = h_t^k \varepsilon_t^{2k}$. Then, as $\mathsf{E}\left[\varepsilon_t^{2k}\right] < \infty$ by assumption, $\mathsf{E}\left[y_t^{2k}\right] < \infty$ if, and only if, $\mathsf{E}\left[h_t^k\right] < \infty$. The binomial formula yields

$$h_t^k = \sum_{p=0}^{k}\binom{k}{p}\left(\alpha_0 + \sum_{i=1}^{H}\alpha_i f_{i,t-1}\right)^p \left[\beta_0 + \lambda_0\varepsilon_{t-1}^2 + \sum_{i=1}^{H}\left(\beta_i + \lambda_i\varepsilon_{t-1}^2\right)f_{i,t-1}\right]^{k-p} h_{t-1}^{k-p}.$$

Let $\mathbf{u}_t = \left[h_t^k, h_t^{k-1}, \ldots, h_t\right]'$. Then,

(A.4) $$\mathbf{u}_t = \mathbf{c}_{t-1} + \mathbf{C}_{t-1}\mathbf{u}_{t-1},$$

where $\mathbf{c}_{t-1} \in \mathbb{R}^k$ is a vector with typical element given by $c_{t-1,j} = \left(\alpha_0 + \sum_{i=1}^{H}\alpha_i f_{i,t-1}\right)^j$, $j = 0, \ldots, k$, and $\mathbf{C}_{t-1}$ is a $k \times k$ upper triangular matrix with diagonal elements given by

$$C_{t-1,jj} = \left[\beta_0 + \lambda_0\varepsilon_{t-1}^2 + \sum_{i=1}^{H}\left(\beta_i + \lambda_i\varepsilon_{t-1}^2\right)f_{i,t-1}\right]^j,$$

$j = 0, \ldots, k$. Therefore,

(A.5) $$\mathsf{E}\left[\mathbf{u}_t|\mathcal{F}_{t-2}\right] \leq const. + \mathsf{E}\left[\mathbf{C}_{t-1}|\mathcal{F}_{t-2}\right]\mathbf{u}_{t-1},$$

since $c_{t-1,j} \leq \left(\alpha_0 + \sum_{i=1}^{H}|\alpha_i|\right)^j < \infty$. $\mathcal{F}_{t-2}$ is the filtration given by all information up to time $t-2$. Because $\mathbf{C}_{t-1}$ is upper triangular, the eigenvalues are equal to the diagonal elements. Thus, the condition

for the existence of the $2k$th-order moment of $y_t$ is

$$(A.6) \quad \mathsf{E}\left[C_{t-1,kk}\right] = \mathsf{E}\left\{\mathsf{E}\left[C_{t-1,kk}|\mathcal{F}_{t-2}\right]\right\} = \mathsf{E}\left\{\left[\beta_0 + \lambda_0\varepsilon_{t-1}^2 + \sum_{i=1}^{H}\left(\beta_i + \lambda_i\varepsilon_{t-1}^2\right)f_{i,t-1}\right]^k\right\} < 1.$$

*Q.E.D*

A.4. **Proof of Corollary 2.** According to results in Theorem 2, the second-order moment condition is

$$(A.7) \qquad\qquad \mathsf{E}\left[\beta_0 + \lambda_0\varepsilon_{t-1}^2 + \sum_{i=1}^{H}\left(\beta_i + \lambda_i\varepsilon_{t-1}^2\right)f_{i,t-1}\right] < 1.$$

Defining two constants $M$ and $N$ and following the same rationale as in the proof of Corollary 1, the result is straightforward. The proof of the fourth-order moment condition follows the same lines and is omitted in order to save space.

This completes the proof.

*Q.E.D*

A.5. **Proof of Theorem 3.** It is easy to see that $G\left(\mathbf{w}_t; \boldsymbol{\psi}\right)$ in (1) is continuous in the parameter vector $\boldsymbol{\psi}$. This follows from the fact that, for each value of $\mathbf{w}_t$, $f\left(y_{t-1}; \gamma_i, c_i\right)$, $i = 1, \ldots, H$, in (1) depend continuously on $\gamma_i$ and $c_i$. Similarly, we can see that $G\left(\mathbf{w}_t, \boldsymbol{\psi}\right)$ is continuous in $\mathbf{w}_t$, and therefore measurable, for each fixed value of the parameter vector $\boldsymbol{\psi}$.

Furthermore, under the restrictions in Assumption 5 and if the stationarity condition of Theorem 1 is satisfied, then $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|h_{u,t}|\right] < \infty$. By Jensen´s inequality $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|\ln|h_{u,t}||\right] < \infty$. Thus $\mathsf{E}\left[|l_{u,t}(\boldsymbol{\psi})|\right] < \infty$, $\forall\,\boldsymbol{\psi}\in\boldsymbol{\Psi}$. This completes the proof.

*Q.E.D*

A.6. **Proof of Theorem 4.** Define $\mathbf{z}_t = \left[1, h_{t-1}, y_{t-1}^2\right]'$ and $\boldsymbol{\phi}_j = \left[\alpha_j, \beta_j, \lambda_j\right]'$, $j = 0, \ldots, H$. Remember also the definition of $\boldsymbol{\theta}_i$ and $\varphi(y_{t-1}; \boldsymbol{\theta}_i)$, $i = 1, \ldots, H$. The parameter vector $\boldsymbol{\psi}$ can be written as $\boldsymbol{\psi} = \left[\boldsymbol{\phi}_0', \ldots, \boldsymbol{\phi}_H', \boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_H'\right]'$. Suppose that $\widetilde{\boldsymbol{\psi}}$ is another vector of parameters such that

$$(A.8) \qquad\qquad \boldsymbol{\phi}_0'\mathbf{z}_t + \sum_{i=1}^{H}\boldsymbol{\phi}_i'\mathbf{z}_t f(y_{t-1}; \boldsymbol{\theta}_i) = \widetilde{\boldsymbol{\phi}}_0'\mathbf{z}_t + \sum_{i=1}^{H}\widetilde{\boldsymbol{\phi}}_i'\mathbf{z}_t f(y_{t-1}; \widetilde{\boldsymbol{\theta}}_i).$$

In order to show global identifiability of the FCGARCH model, we need to prove that, under Restrictions (R.1) and (R.2) in Assumptions 3 and 6, (A.8) is satisfied if, and only if, $\boldsymbol{\psi} = \widetilde{\boldsymbol{\psi}}$.

Equation (A.8) can be rewritten as

$$(A.9) \qquad\qquad \boldsymbol{\phi}_0'\mathbf{z}_t - \widetilde{\boldsymbol{\phi}}_0'\mathbf{z}_t - \sum_{j=1}^{2H}\overline{\boldsymbol{\phi}}_j'\mathbf{z}_t f\left(y_{t-1}; \overline{\boldsymbol{\theta}}_j\right) = 0,$$

where $\overline{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_j$ for $j = 1, \ldots, H$, $\overline{\boldsymbol{\theta}}_j = \widetilde{\boldsymbol{\theta}}_{j-H}$ for $j = H+1, \ldots, 2H$, $\overline{\boldsymbol{\phi}}_j = \boldsymbol{\phi}_j$ for $j = 1, \ldots, H$, and $\overline{\boldsymbol{\phi}}_j = \boldsymbol{\phi}_{j-H}$ for $j = H+1, \ldots, 2H$.

For simplicity of notation let $\varphi_j = \varphi\left(y_{t-1}; \boldsymbol{\theta}_j\right)$, $j = 1, \ldots, 2H$. Lemma 2.7 in Hwang and Ding (1997) implies that if $\varphi_{j_1}$ and $\varphi_{j_2}$ are not sign-equivalent, $j_1 \in \{1, \ldots, 2H\}$, $j_2 \in \{1, \ldots, 2H\}$, (A.9) holds if, and only if, $\boldsymbol{\phi}_0$, $\widetilde{\boldsymbol{\phi}}_0$, and $\overline{\boldsymbol{\phi}}_j$ vanish jointly for every $j \in \{1, \ldots, 2H\}$. However, Assumption 6 precludes

that possibility. Hence, $\varphi_{j_1}$ and $\varphi_{j_2}$ must be sign-equivalent. But restriction (R.2) in Assumption 4 avoid that two functions $\varphi_{j_1}$ and $\varphi_{j_2}$ coming from the same model being sign-equivalent. Consequently, $\exists\, j_1 \in \{0, \ldots, H\}$ and $j_2 \in \{H{+}1, \ldots, 2H{+}1\}$ such that $\varphi_{j_1}$ and $\varphi_{j_2}$ are sign-equivalent. Under restriction (R.2) the only possibility is that the regimes are permuted. Restriction (R.1) excludes that possibility. Hence, the only case where (A.8) holds is when $\phi_i = \widetilde{\phi}_i$, and $\boldsymbol{\theta}_i = \widetilde{\boldsymbol{\theta}}_i$, $i = 1, \ldots, H$.

Let $h_t^{(0)}$ be the true conditional variance. To show that $\mathcal{L}(\boldsymbol{\psi})$ is uniquely maximized at $\boldsymbol{\psi}_0$, rewrite the maximization problem as

$$\max_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} [\mathcal{L}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi}_0)] = \max_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left\{ -\mathsf{E}\left[ -\ln\left(\frac{h_t^{(0)}}{h_{u,t}}\right) + \frac{h_t^{(0)}}{h_{u,t}} + 1 \right] \right\}.$$

In addition, for any $x > 0$, $-m(x) = -\ln(x) + x \geq 0$, so that

$$-\mathsf{E}\left[ -\ln\left(\frac{h_t^{(0)}}{h_{u,t}}\right) + \frac{h_t^{(0)}}{h_{u,t}} \right] \leq 0.$$

Furthermore, $m(x)$ is maximized at $x = 1$. If $x \neq 1$, $m(x) < m(1)$, implying that $\mathsf{E}[m(x)] \leq \mathsf{E}[m(1)]$ with equality only if $x = 1$ with probability one. However, this will occur only if $-\ln\left(\frac{h_t^{(0)}}{h_{u,t}}\right) = 0$ with probability one. By the mean value theorem, this is equivalent to showing that

$$(\boldsymbol{\psi} - \boldsymbol{\psi}_0)\frac{\partial h_{u,t}}{\partial \boldsymbol{\psi}}\frac{1}{h_{u,t}} = 0$$

with probability one. By Lemma 1 this occurs if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. This completes the proof.

*Q.E.D*

A.7. **Proof of Theorem 5.** Following Newey and McFadden (1994), [7] $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{p} \boldsymbol{\psi}_0$ if the following conditions hold:

(1) The parameter space $\boldsymbol{\Psi}$ is compact.
(2) $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. Furthermore $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is a measurable function of $y_t$, $t = 1, \ldots, T$, for all $\boldsymbol{\psi} \in \boldsymbol{\Psi}$.
(3) $\mathcal{L}(\boldsymbol{\psi})$ has a unique maximum at $\boldsymbol{\psi}_0$.
(4) $\mathcal{L}_{u,T}(\boldsymbol{\psi}) \xrightarrow{p} \mathcal{L}(\boldsymbol{\psi})$.

Condition (1) is met by assumption. Theorem 3 shows that Condition (2) is trivially satisfied. Theorem 4 proves that Condition (3) is fulfilled and, by Lemma 2, Condition (4) is also satisfied. Thus, $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{p} \boldsymbol{\psi}_0$.

Lemma 3 shows that

$$\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})| \xrightarrow{p} 0,$$

implying that $\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_0$. This completes the proof.

*Q.E.D*

A.8. **Proof of Theorem 6.** We start proving asymptotically normality of the QMLE using the unobserved log-likelihood. Once this is shown, the proof using the observed log-likelihood is immediate by Lemmas 3 and 5. To prove the asymptotically normality of the QMLE we need the following conditions in addition to the ones stated in the proof of Theorem 5.

---

[7]See also White (1994) and Wooldridge (1994).

(5) The true parameter vector $\boldsymbol{\psi}_0$ is interior to $\boldsymbol{\Psi}$.

(6) The matrix

$$\mathbf{A}_T(\boldsymbol{\psi}) = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial^2 l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right)$$

exists and is continuous in $\boldsymbol{\Psi}$.

(7) The matrix $\mathbf{A}_T(\boldsymbol{\psi}) \overset{p}{\to} \mathbf{A}(\boldsymbol{\psi}_0)$, for any sequence $\boldsymbol{\psi}_T$ such that $\boldsymbol{\psi}_T \overset{p}{\to} \boldsymbol{\psi}_0$.

(8) The score vector satisfies

$$\frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right) \overset{D}{\to} \mathsf{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\psi}_0)).$$

Condition (5) is satisfied by assumption. Condition (6) follows from the fact that $l_t(\boldsymbol{\psi})$ is differentiable of order two on $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ and the stationarity of the FCGARCH model. Lemma 5 imply that Condition (7) is satisfied. Furthermore, non-singularity of $\mathbf{A}(\boldsymbol{\psi}_0)$ follows immediately from identification of the FCGARCH model and the non-singularity of $\mathbf{B}(\boldsymbol{\psi}_0)$; see Hwang and Ding (1997). In Lemma 4 we prove that condition (8) is also met. This completes the proof.

*Q.E.D*

A.9. **Proof of Theorem 7.** The local approximation to the instantaneous quasi-log-likelihood function in a neighborhood of $\mathcal{H}_0$ is

(A.10) $$l_t = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln (h_t) - \frac{y_t^2}{2h_t},$$

with $h_t$ given by (17). Let $\boldsymbol{\psi} = \left[ \boldsymbol{\psi}_1', \boldsymbol{\psi}_2' \right]'$ with

$$\boldsymbol{\psi}_1 = \left[ \tilde{\alpha}_0, \tilde{\beta}_0, \tilde{\lambda}_0, \alpha_1, \ldots, \alpha_{H-1}, \beta_1, \ldots, \beta_{H-1}, \lambda_1, \ldots, \lambda_{H-1}, \gamma_1, \ldots, \gamma_{H-1}, c_1, \ldots, c_{H-1} \right]'$$

and $\boldsymbol{\psi}_2 = [\pi, \delta, \rho]'$.

Furthermore, it can be shown that the score vector is given by

(A.11) $$\mathbf{q}(\boldsymbol{\psi}) = [\mathbf{q}(\boldsymbol{\psi}_1)', \mathbf{q}(\boldsymbol{\psi}_2)']' = \sum_{t=1}^{T} \begin{bmatrix} \frac{\partial l_t}{\partial \boldsymbol{\psi}_1} \\ \frac{\partial l_t}{\partial \boldsymbol{\psi}_2} \end{bmatrix} = \sum_{t=1}^{T} \frac{1}{2} \left( \frac{y_t^2}{h_t} - 1 \right) \begin{bmatrix} \mathbf{z}_t \\ \mathbf{u}_t \end{bmatrix}$$

where $\mathbf{z}_t = \frac{1}{h_t} \frac{\partial h_t}{\partial \boldsymbol{\psi}_1}$ and $\mathbf{u}_t = \frac{1}{h_t} \frac{\partial h_t}{\partial \boldsymbol{\psi}_2}$.

The information matrix is given by

$$\mathbf{A}(\boldsymbol{\psi}) = \mathsf{E}\left[ -\frac{\partial^2 l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right] = \mathsf{E}\left[ \frac{1}{2h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\psi}} \frac{\partial h_t}{\partial \boldsymbol{\psi}'} \left( \frac{y_t^2}{h_t} \right) - \left( \frac{y_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \boldsymbol{\psi}'} \left( \frac{1}{2h_t} \frac{\partial h_t}{\partial \boldsymbol{\psi}} \right) \right]$$

$$= \mathsf{E}\left[ \frac{1}{2h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\psi}} \frac{\partial h_t}{\partial \boldsymbol{\psi}'} \right] = \frac{1}{2} \mathsf{E}\begin{bmatrix} \mathbf{z}_t \mathbf{z}_t' & \mathbf{z}_t \mathbf{u}_t' \\ \mathbf{u}_t \mathbf{z}_t' & \mathbf{u}_t \mathbf{u}_t' \end{bmatrix}.$$

A consistent estimator for $\mathbf{A}(\boldsymbol{\psi})$ is

(A.12) $$\mathbf{A}_T(\boldsymbol{\psi}) = \frac{1}{2T} \sum_{t=1}^{T} \begin{bmatrix} \mathbf{z}_t \mathbf{z}_t' & \mathbf{z}_t \mathbf{u}_t' \\ \mathbf{u}_t \mathbf{z}_t' & \mathbf{u}_t \mathbf{u}_t' \end{bmatrix};$$

see Engle (1982).

Defining $\mathbf{d}_t = [\mathbf{z}_t', \mathbf{u}_t']$ and following Godfrey (1988, p. 16), the $LM$ statistic is given by

$$LM = \mathbf{q}(\boldsymbol{\psi})|_{\mathcal{H}_0} \left[\mathbf{A}_T(\boldsymbol{\psi})|_{\mathcal{H}_0}\right]^{-1} \mathbf{q}(\boldsymbol{\psi})|_{\mathcal{H}_0}$$

(A.13)
$$= \frac{T}{2} \sum_{t=1}^{T} \left(\frac{y_t^2}{h_t} - 1\right) \mathbf{d}_t \left[\sum_{t=1}^{T} \mathbf{d}_t \mathbf{d}_t'\right]^{-1} \sum_{t=1}^{T} \left(\frac{y_t^2}{h_t} - 1\right) \mathbf{d}_t.$$

Then, Lemmas 4 and 5 yield the result. This completes the proof.

<div align="right">*Q.E.D*</div>

## APPENDIX B. LEMMAS

LEMMA 1. *Suppose that $y_t$ is generated by (1), satisfying Assumptions 3–5. Let $\mathbf{d}$ be a constant vector with the same dimension as $\boldsymbol{\psi}$. Then*

$$\mathbf{d}' \left(\frac{\partial h_{u,t}}{\partial \boldsymbol{\psi}}\right) = 0 \text{ a.s. iff } \mathbf{d} = \mathbf{0}.$$

PROOF. The proof follows the same reasoning as the one from Lemma 5 in Lumsdaine (1996). Define $\boldsymbol{\xi}_t = \frac{\partial h_t}{\partial \boldsymbol{\psi}}$ and $f_{i,t} \equiv f(y_t; \gamma_i, c_i)$. It is straightforward to show that

$$\boldsymbol{\xi}_t = \overline{\beta}(y_{t-1})\boldsymbol{\xi}_{t-1} + \boldsymbol{\kappa}_{t-1},$$

where

$$\begin{aligned}
\boldsymbol{\kappa}_{t-1} = \Big[ & 1, \, h_{t-1}, \, y_{t-1}^2, \, f_{1,t-1}, \, \ldots, \, f_{H,t-1}, \\
& f_{1,t-1}h_{t-1}, \, \ldots, \, f_{H,t-1}h_{t-1}, \, f_{1,t-1}y_{t-1}^2, \, \ldots, \, f_{H,t-1}y_{t-1}^2, \\
& \left(\alpha_1 + \beta_1 h_{t-1} + \lambda_1 y_{t-1}^2\right) \frac{\partial f_{1,t-1}}{\partial \gamma_1}, \, \ldots, \, \left(\alpha_H + \beta_H h_{t-1} + \lambda_H y_{t-1}^2\right) \frac{\partial f_{H,t-1}}{\partial \gamma_H}, \\
& \left(\alpha_1 + \beta_1 h_{t-1} + \lambda_1 y_{t-1}^2\right) \frac{\partial f_{1,t-1}}{\partial c_1}, \, \ldots, \, \left(\alpha_H + \beta_H h_{t-1} + \lambda_H y_{t-1}^2\right) \frac{\partial f_{H,t-1}}{\partial c_H} \Big]',
\end{aligned}$$

and $\overline{\beta}(y_t) = \beta_0 + \sum_{i=1}^{H} \beta_i f_{i,t}$. Then, $\mathbf{d}'\boldsymbol{\xi}_t = \mathbf{d}'\overline{\beta}(y_{t-1})\boldsymbol{\xi}_{t-1} + \mathbf{d}'\boldsymbol{\kappa}_{t-1}$. Since by assumption $\mathbf{d}'\boldsymbol{\xi}_t = 0$ and $\mathbf{d}'\boldsymbol{\xi}_{t-1} = 0$ with probability one, this implies that $\mathbf{d}'\boldsymbol{\kappa}_{t-1} = 0$ with probability one. Since $\boldsymbol{\kappa}_t$ is nondegenarate, $\mathbf{d}'\boldsymbol{\xi}_t = 0$ with probability one if and only if $\mathbf{d} = \mathbf{0}$. This completes the proof.

<div align="right">*Q.E.D*</div>

LEMMA 2. *Under assumptions of Lemma 1,*

$$\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi})| \xrightarrow{p} 0.$$

PROOF. The proof of this lemma follows closely the proof of Lemma 4.3 in Ling and McAleer (2003). Set $g(\mathbf{Y}_t, \boldsymbol{\psi}) = l_{u,t}(\boldsymbol{\psi}) - \mathsf{E}\left[l_{u,t}(\boldsymbol{\psi})\right]$, where $\mathbf{Y}_t = [y_t, y_{t-1}, y_{t-2}, \ldots]'$. Theorem 3 implies that $\mathsf{E}\left[\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |g(\mathbf{Y}_t, \boldsymbol{\psi})|\right] < \infty$. In addition, because $g(\mathbf{Y}_t, \boldsymbol{\psi})$ is stationary with $\mathsf{E}\left[g(\mathbf{Y}_t, \boldsymbol{\psi})\right] = 0$, by Theorem 3.1 in Ling and McAleer (2003) $\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left|T^{-1} \sum_{t=1}^{T} g(\mathbf{Y}_t, \boldsymbol{\psi})\right| = o_p(1)$. This completes the proof.

<div align="right">*Q.E.D*</div>

LEMMA 3. *Under assumptions of Lemma 1*

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} |\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})| \xrightarrow{p} 0.$$

PROOF. The proof of this lemma is adapted from the proof of item (i) of Lemma 6 in Lumsdaine (1996). Write

$$h_{u,0} = \sum_{k=1}^{\infty} \phi_{-k}(\boldsymbol{\psi}) + \sum_{k=1}^{\infty} \widetilde{\phi}_{-k}(\boldsymbol{\psi}) y_{-k}^2.$$

It is clear that under the condition of Theorem 1,

$$\Pr\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} (h_{u,0}) > K\right] \to 0 \text{ as } K \to \infty$$

and $h_{u,0}$ is well defined.

Set $\overline{\beta}(y_t) = \beta_0 + \sum_{i=1}^H \beta_i f(y_t; \gamma_i, c_i)$. Combining (9) and (11) we get

$$h_{u,t} - h_t = \left(\prod_{j=1}^t \overline{\beta}(y_j)\right)(h_{u,0} - h_0) \text{ or } h_{u,t} + \left(\prod_{j=1}^t \overline{\beta}(y_j)\right) h_0 = h_t - \left(\prod_{j=1}^t \overline{\beta}(y_j)\right) h_{u,0}.$$

Therefore, dividing the left-hand-side of the above expression by $h_t$, we have:

$$\left|\ln\left(\frac{h_{u,t}}{h_t}\right)\right| < \left|\ln\left[1 + \left(\prod_{j=1}^t \overline{\beta}(y_j)\right)\frac{h_0}{h_t}\right]\right| < \left(\prod_{j=1}^t \overline{\beta}(y_j)\right)\frac{h_{u,0}}{h_t}.$$

Define two finite positive constants $\underline{\delta}$ and $\overline{\delta}$. From the fact that $h_t > \underline{\delta}$ and $\overline{\beta}(y_t) \leq \overline{\delta}$

$$0 \leq \left|T^{-1/2}\sum_{t=1}^T \ln\left(\frac{h_{u,t}}{h_t}\right)\right|^p \leq \left[T^{-1/2}\sum_{t=1}^T \left|\ln\left(\frac{h_{u,t}}{h_t}\right)\right|\right]^p \leq \left[T^{-1/2}\sum_{t=1}^T \left(\prod_{j=1}^t \overline{\beta}(y_j)\right)\frac{h_{u,0}}{h_t}\right]^p$$

$$\leq T^{-p/2}\frac{h_{u,0}^p}{\underline{\delta}^p}\left[\sum_{t=1}^T \left(\prod_{j=1}^t \overline{\beta}(y_j)\right)\right]^p.$$

Then, by Theorem 1 and Slutsky´s Theorem, the upper bound of the final expression converges in probability uniformly to zero, so that

$$\Pr\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\sum_{t=1}^T |\ln(h_{u,t}) - \ln(h_t)| > K\right] \to 0 \text{ as } T \to \infty, \forall K > 0.$$

Now we have to prove that

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|T^{-1/2}\sum_{t=1}^T \left(\frac{y_t^2}{h_{u,t}} - \frac{y_t^2}{h_t}\right)\right| \xrightarrow{p} 0.$$

It is clear that

$$\left[T^{-1/2}\sum_{t=1}^T \left|y_t^2\left(\frac{h_t - h_{u,t}}{h_{u,t}h_t}\right)\right|\right]^p \leq \frac{1}{T^{p/2}\underline{\delta}^{2p}}\left[\sum_{t=1}^T |y_t^2(h_t - h_{u,t})|\right]^p$$

$$= \frac{1}{T^{p/2}\underline{\delta}^{2p}}\left[\sum_{t=1}^T y_t^2 \left|\left(\prod_{j=1}^t \overline{\beta}(y_j)\right)(h_0 - h_{u,0})\right|\right]^p.$$

Define $X_t \equiv y_t^2$ and $\xi_t \equiv \left| \left( \prod_{j=1}^t \overline{\beta}(y_j) \right) (h_0 - h_{u,0}) \right|$. Under the condition of Theorem 1 $X_t$ is a strictly stationary and ergodic time series, with $\mathsf{E}\left[ \| X_t \| \right] < \infty$. Furthermore, it is clear by Theorem 1 that $\sup_t |\xi_t| \le C$, where $C$ is a finite constant and $T^{-1} \sum_{t=1}^T |\xi_t| = o_p(1)$. Then, by Lemma 4.5 in Ling and McAleer (2003) $T^{-1} \sum_{t=1}^T X_t \xi_t = o_p(1)$. Hence,

$$\frac{1}{T^{p/2} \underline{\delta}^{2p}} \left[ \sum_{t=1}^T y_t^2 \left| \left( \prod_{j=1}^t \overline{\beta}(y_j) \right) (h_0 - h_{u,0}) \right| \right]^p \xrightarrow{p} 0.$$

This completes the proof.

*Q.E.D*

LEMMA 4. *Under the conditions of Theorem 6, then* $\mathsf{E}\left[ \left. \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \right]$ *exists and is finite. Furthermore,* $\mathbf{B}(\boldsymbol{\psi}_0)$ *is finite, positive definite, and*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left. \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \xrightarrow{D} \mathsf{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\psi}_0)).$$

PROOF. First, it is straightforward to show that, if the condition of Theorem 1 is satisfied, then $\mathsf{E}\left[ \left. \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \right]$ exists and is finite. Now, set

(B.14) $$\nabla_0 l_{u,t} \equiv \left. \frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}=\boldsymbol{\psi_0}} \quad \text{and} \quad \nabla_0 h_{u,t} \equiv \left. \frac{\partial h_{u,t}}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}=\boldsymbol{\psi_0}}.$$

Then,

(B.15) $$\nabla_0 l_{u,t} \nabla_0 l'_{u,t} = \frac{1}{4} \left( \varepsilon_t^4 - 2\varepsilon_t^2 + 1 \right) \frac{1}{h_{u,t}^2} \nabla_0 h_{u,t} \nabla_0 h'_{u,t}.$$

Let $\underline{\delta} < \infty$ be a positive constant such that $h_{u,t} > \underline{\delta}$. If the log-moment condition of Theorem 1 is met, then, using the same arguments as in the proof of Lemma 1 in Boussama (2000),

$$\mathsf{E}\left[ \frac{1}{h_{u,t}^2} \nabla_0 h_{u,t} \nabla_0 h'_{u,t} \right] < \mathbf{M},$$

where $\mathbf{M}$ is a constant vector with finite elements, and

(B.16) $$\mathsf{E}\left[ \nabla_0 l_{u,t} \nabla_0 l'_{u,t} \right] \le \frac{1}{4} \mathbf{M} \mathsf{E}\left[ (\varepsilon_t^4 - 2\varepsilon_t^2 + 1) \right] = \frac{1}{4} \mathbf{M} \left( \mu_4 - 1 \right),$$

where $\mu_4 = \mathsf{E}\left[ \varepsilon_t^4 \right] < \infty$. Hence, $\mathsf{E}\left[ \nabla_0 l_{u,t} \nabla_0 l'_{u,t} \right] \le \infty$ and $\mathbf{B}(\boldsymbol{\psi}_0) < \infty$. Under conditions of Theorems 1 and 4, $\mathbf{B}(\boldsymbol{\psi}_0)$ is positive definite.

Now, let $S_T = \sum_{t=1}^T \mathbf{c}' \nabla_0 l_{u,t}$, where $\mathbf{c}$ is a constant vector. Then $S_T$ is a martingale with respect to $\mathcal{F}_t$, the filtration generated by all past observations of $y_t$. By the given assumptions $\mathsf{E}\left[ S_T \right] > 0$. Using the central limit theorem of Stout (1974)

$$T^{-1/2} S_T \xrightarrow{D} \mathsf{N}\left( 0, \mathbf{c}' \mathbf{B}(\boldsymbol{\psi}_0) \mathbf{c} \right).$$

Finally, by the Cramér-Wold device,

$$T^{-1/2} \sum_{t=1}^{T} \left. \frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi} = \boldsymbol{\psi_0}} \xrightarrow{D} \mathsf{N}\left(0, \mathbf{B}(\boldsymbol{\psi}_0)\right).$$

In a similar manner to the proof of Lemma 3, we can show that

$$T^{-1/2} \sum_{t=1}^{T} \left\| \left. \frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi} = \boldsymbol{\psi_0}} - \left. \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi} = \boldsymbol{\psi_0}} \right\| \xrightarrow{p} \mathbf{0}.$$

Thus

$$T^{-1/2} \sum_{t=1}^{T} \left. \frac{\partial l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi} = \boldsymbol{\psi_0}} \xrightarrow{D} \mathsf{N}(0, \boldsymbol{B}_0).$$

This completes the proof.

<div align="right">*Q.E.D*</div>

LEMMA 5. *Under the conditions of Theorem 6*

(a) $$\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} - \mathsf{E}\left[ \frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right] \right\| \xrightarrow{p} \mathbf{0}, \text{ and}$$

(b) $$\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left\| \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} - \frac{\partial^2 l_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right] \right\| \xrightarrow{p} \mathbf{0}.$$

PROOF. First set

(B.17) $$\nabla_0^2 l_{u,t} \equiv \left. \frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right|_{\boldsymbol{\psi} = \boldsymbol{\psi_0}}, \text{ and } \nabla_0^2 h_{u,t} \equiv \left. \frac{\partial^2 h_{u,t}}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right|_{\boldsymbol{\psi} = \boldsymbol{\psi_0}}.$$

Then,

(B.18) $$\nabla_0^2 l_{u,t} = \left( \frac{y_t^2}{h_{u,t}} - 1 \right) \frac{1}{2 h_{u,t}} \nabla_0^2 h_{u,t} - \frac{1}{2 h_{u,t}^2} \left( \frac{y_t^2}{h_{u,t}} \right) \nabla_0 h_{u,t} \nabla_0 h_{u,t}'$$

Under the condition of Theorem 1 it can be shown, as $\nabla_0^2 h_{u,t}$ has only second-order terms, that $\mathsf{E}\left[ \nabla_0^2 h_{u,t} \right] \leq \mathbf{M}_2 < \infty$, where $\mathbf{M}_2$ is a constant matrix. Then,

(B.19) $$\nabla_0^2 l_{u,t} \leq \left( \varepsilon_t^2 - 1 \right) \frac{1}{\underline{\delta}} \mathbf{M}_2 - \frac{1}{2 \underline{\delta}^2} \varepsilon_t^2 \mathbf{M}$$

and $\mathsf{E}\left[ \nabla_0^2 l_{u,t} \right] \leq \infty$. By Theorem 3.1 in Ling and McAleer (2003), (a) holds. The proof of (b) follows closely the one of Lemma 3 and the details are omitted.

This completes the proof.

<div align="right">*Q.E.D*</div>

## REFERENCES

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001a): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2001b): "Exchange rate returns standardized by realized volatility are (nearly) Gaussian," *Multinational Finance Journal*, forthcoming.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

ANDERSON, H. M., K. NAM, AND F. VAHID (1999): "Asymmetric Nonlinear Smooth Transition GARCH Models," in *Nonlinear Time Series Analysis of Economic and Financial Data*, ed. by P. Rothman, pp. 191–207. Kluwer.

AUDRINO, F., AND P. BÜHLMANN (2001): "Tree-Structured GARCH Models," *Journal of the Royal Statistical Society, Series B*, 63, 727–744.

AUDRINO, F., AND F. TROJANI (forthcoming): "Estimating and Predicting Multivariate Volatility Regimes in Global Stock Markets," *Journal of Applied Econometrics*.

BACON, D. W., AND D. G. WATTS (1971): "Estimating the Transition Between Two Intersecting Lines," *Biometrika*, 58, 525–534.

BERKES, I., L. HORVÁTH, AND P. KOKOSZKA (2003): "GARCH Processes: Structure and Estimation," *Bernoulli*, 9, 201–227.

BLACK, F. (1976): "Studies in Stock Price Volatility Changes," *Proceedings of The American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 21, 307–328.

BOUSSAMA, F. (2000): "Normalité Asymptotique de L´Estimateur du Pseudo-Maximum de Vraisemblance dún Modèle GARCH," *Comptes Rendus de l´Academie des Sciences, Série I*, 331, 81–84.

CAI, J. (1994): "A Markov Model of Switching-Regime ARCH," *Journal of Business and Economic Statistics*, 12, 309–316.

CARNERO, M. A., D. PEÑA, AND E. RUIZ (2004): "Persistence and kurtosis in GARCH and stochastic volatility models," *Journal of Financial Econometrics*, 2, 319–342.

CARRASCO, M., AND X. CHEN (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 17–39.

CHAN, K. S., AND H. TONG (1986): "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis*, 7, 179–190.

CHEN, C., T. CHIANG, AND M. SO (2003): "Asymmetrical Reaction to US Stock-Return News: Evidence from Major Stock Markets Based on a Double-Threshold Model," *Journal of Economics and Business*, 55, 487–502.

CHRISTOFFERSEN, P. F. (1998): "Evaluating interval forecasts," *International Economic Review*, 39, 841–862.

COMTE, F., AND O. LIEBERMAN (2003): "Asymptotic Theory for Multivariate GARCH Processes," *Journal of Multivariate Analysis*, 84, 61–84.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long MemoryProperty of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflations," *Econometrica*, 50, 987–1007.

ENGLE, R. F., AND V. K. NG (1993): "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

FORNARI, F., AND A. MELE (1997): "Sign- and Volatility-Switching ARCH Models: Theory and Applications to International Stock Markets," *Journal of Applied Econometrics*, 12, 49–65.

FRANCQ, C., AND J.-M. ZAKOÏAN (2004): "Maximum Likelihood Estimatin of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637.

GLOSTEN, L., R. JAGANNANTHAN, AND R. RUNKLE (1993): "On The Relationship Between The Expected Value and The Volatility of The Nominal Excess Returns on Stocks," *Journal of Finance*, 48, 1779–1801.

GODFREY, L. G. (1988): *Misspecification Tests in Econometrics*, vol. 16 of *Econometric Society Monographs*. Cambridge University Press, New York, NY, second edn.

GOETZMANN, W. N., R. G. IBBOTSON, AND L. PENG (2001): "A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability," *Journal of Financial Markets*, 4, 1–32.

GONZALEZ-RIVERA, G. (1998): "Smooth Transition GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 3(2), 61–78.

GOURIEROUX, C., AND A. MONFORT (1995): *Statistics and Econometric Models*, vol. 2. Cambridge University Press, Melbourne.

GRANGER, C., AND Z. DING (1995): "Some Properties of Absolute Returns: An Alternative Measure of Risk," *Annales d'Economie et de Statistique*, 40, 67–95.

HAGERUD, G. E. (1997): "A New Non-Linear GARCH Model," Phd. thesis, Stockholm School of Economics.

HAMILTON, J. D., AND R. SUSMEL (1994): "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307–333.

HANSEN, B. E. (1996): "Inference when a Nuisance Parameter is not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

HE, C., AND T. TERÄSVIRTA (1999): "Properties of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 92, 173–192.

HWANG, J. T. G., AND A. A. DING (1997): "Prediction Intervals for Artificial Neural Networks," *Journal of the American Statistical Association*, 92, 109–125.

JEANTHEAU, T. (1998): "Strong Consistency of Estimators for Multivariate ARCH Models," *Econometric Theory*, 14, 70–86.

JENSEN, S. T., AND A. RAHBEK (2004): "Asymptotic Normality of the QMLE Estimator of ARCH in the Nonstationary Case," *Econometrica*, 72, 641–646.

LANNE, M., AND P. SAIKKONEN (2005): "Nonlinear GARCH Models for Highly Persistent Volatility," *Econometrics Journal*, 8, 251–276.

LEE, S.-W., AND B. E. HANSEN (1994): "Asymptotic Theory for The GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

LI, C. W., AND W. K. LI (1996): "On a Double Threshold Autoregressive Heteroscedastic Time Series Model," *Journal of Applied Econometrics*, 11, 253–274.

LI, W. K., S. LING, AND M. MCALEER (2002): "Recent Theoretical Results for Time Series Models with GARCH Errors," *Journal of Economic Surveys*, 16, 245–269.

LING, S., AND M. MCALEER (2002): "Stationarity and The Existence of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 106, 109–117.

——— (2003): "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

LIU, J., W. K. LI, AND C. W. LI (1997): "On a Threshold Autoregression with Conditional Heteroscedastic Variances," *Journal of Statistical Planning and Inference*, 62(2), 279–300.

LUMSDAINE, R. (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.

LUNDBERGH, S., AND T. TERÄSVIRTA (2002): "Evaluating GARCH Models," *Journal of Econometrics*, 110, 417–435.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

MALMSTEN, H., AND T. TERÄSVIRTA (2004): "Stylized Facts of Financial Time Series and Three Popular Models of Volatility," Working Paper Series in Economics and Finance 563, Stockholm School of Economics.

McAleer, M. (2005): "Automated Inference and Learning in Modeling Financial Volatility," *Econometric Theory*, 21, 232–261.

McAleer, M., F. Chan, and D. Marinova (forthcoming): "An Econometric Analysis of Asymmetric Volatility: Theory and Application to Patents," *Journal of Econometrics*.

Medeiros, M. C., T. Teräsvirta, and G. Rech (in press): "Building Neural Network Models for Time Series: A Statistical Approach," *Journal of Forecasting*.

Medeiros, M. C., and A. Veiga (2005): "A Flexible Coefficient Smooth Transition Time Series Model," *IEEE Transactions on Neural Networks*, 16, 97–113.

Nelson, D. B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

——— (1991): "Conditinal Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Newey, W., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2111–2245. Elsevier Science.

Rabemananjara, R., and J. M. Zakoian (1993): "Threshold ARCH Models and Asymmetries in Volatility," *Journal of Applied Econometrics*, 8(1), 31–49.

Schwert, G. W. (1990): "Stock Volatility and the Crash of '87," *Review of Financial Studies*, 3, 77–102.

Stout, W. F. (1974): *Almost Sure Convergence*. Academic Press, New York.

Sussman, H. J. (1992): "Uniqueness of the Weights for Minimal Feedforward Nets with a Given Input-Output Map," *Neural Networks*, 5, 589–593.

Teräsvirta, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218.

Teräsvirta, T., and I. Mellin (1986): "Model Selection Criteria and Model Selection Tests in Regression Models," *Scandinavian Journal of Statistics*, 13, 159–171.

Trapletti, A., F. Leisch, and K. Hornik (2000): "Stationary and Integrated Autoregressive Neural Network Processes," *Neural Computation*, 12, 2427–2450.

van Dijk, D., P. Franses, and A. Lucas (1999a): "Testing for ARCH in the presence of additive outliers," *Journal of Business and Economic Statistics*, 17, 217–235.

——— (1999b): "Testing for smooth transition nonlinearity in the presence of outliers," *Journal of Applied Econometrics*, 14, 539–562.

van Dijk, D., and P. H. Franses (1999): "Modelling Multiple Regimes in the Business Cycle," *Macroeconomic Dynamics*, 3(3), 311–340.

White, H. (1994): *Estimation, Inference and Specification Analysis*. Cambridge University Press, New York, NY.

——— (2001): *Asymptotic Theory for Econometricians*. Academic Press, San Diego.

Wooldridge, J. M. (1990): "A unified approach to robust, regression-based specification tests," *Econometric Theory*, 6, 17–43.

——— (1994): "Estimation and Inference for Dependent Process," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2639–2738. Elsevier Science.

Zakoian, J. M. (1994): "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18, 931–955.

(M. C. Medeiros) Department of Economics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

*E-mail address*: mcm@econ.puc-rio.br

(A. Veiga) Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

*E-mail address*: alvf@ele.puc-rio.br