Econometric Theory, **25**, 2009, 117–161. Printed in the United States of America.
doi:10.1017/S026646660809004X

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

MARCELO C. MEDEIROS AND ALVARO VEIGA
Pontifical Catholic University of Rio de Janeiro

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 tests that circumvents identification problems that are usually found in nonlinear time series models. The second advantage is that the novel 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 and 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.

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

This research has been partially supported by CNPq and FAPERJ. A previous version of this paper has circulated under the title "Are There Multiple Regimes in Financial Volatility?" Material from this paper has been presented at several conferences. We thank the participants for useful comments and suggestions. Our thanks also go to Marcelo Fernandes, Changli He, Michael McAleer, Walter Novaes, Leonardo Souza, Timo Teräsvirta, and Dick van Dijk for careful reading and useful comments. Part of this work was carried out while the first author was visiting the Department of Economic Statistics at the Stockholm School of Economics and the School of Commerce and Economics at the University of Western Australia. Their kind hospitality is gratefully acknowledged. Finally, we thank the co-editor, Bruce Hansen, and three referees for comments and suggestions that led to an improved version of the paper. The responsibility for any errors or shortcomings in the paper remains ours. Address correspondence to Marcelo C. Medeiros, Department of Economics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil, e-mail: mcm@econ.puc-rio.br.

© 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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 autoregressive conditional heteroskedasticity (ARCH) model proposed by Engle (1982) and the generalized autoregressive conditional heteroskedasticity (GARCH) 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 previously mentioned asymmetry has motivated a large number of different volatility models that have been applied with relative 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 autoregressive moving average (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 Zakoïan (1993) and Zakoïan (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. Although 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 depend on the sign of the past returns. The difference is that the former allows for a smooth transition between regimes. In

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

the ESTGARCH 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, which is able to model both sign and size asymmetries. Liu, Li, and Li (1997) generalized it, proposing the double threshold GARCH (DTGARCH) 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 previously. As most of the empirical papers in the financial econometrics literature deal only with GARCH(1,1)-type models, we focus our attention only on the first-order FCGARCH specification. Furthermore, analyzing higher order FCGARCH models, although relevant to practitioners, may overly complicate the theoretical derivations and will be left for future research.[2]

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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

fail to describe satisfactorily. Reproducing the previously mentioned typical behavior of financial time series may be 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 previously 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 (QMLEs) Consistency and asymptotic normality are proved under weak conditions. Our results are directly applicable to other nonlinear GARCH specifications, such as the STGARCH model. 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 10 stock indexes shows evidence of two regimes for six series and three regimes for the other four series. 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 standard 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 corroborating the hypothesis of both size and sign asymmetries. When a forecasting exercise is considered, the FCGARCH model outperforms the GARCH, GJR, EGARCH, and tree-structured GARCH alternatives for eight out of 10 cases.

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 two Appendixes.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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 FCGARCH model with* $m = H + 1$ *limiting regimes, FCGARCH*(m,1,1)*, if*

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

$$h_t = G(\mathbf{w}_t; \boldsymbol{\psi}) = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \sum_{i=1}^{H} \left[\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2\right] f(s_t; \gamma_i, c_i), \qquad t = 1, \ldots, T, \tag{1}$$

where $\{\varepsilon_t\}$ *is a sequence of identically and independently distributed zero mean and unit variance random variables,* $\varepsilon_t \sim \mathrm{IID}(0,1)$*,* $G(\mathbf{w}_t; \boldsymbol{\psi})$ *is a nonlinear function of the vector of variables* $\mathbf{w}_t = \left[y_{t-1}, h_{t-1}, s_t\right]'$ *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 $f(s_t; \gamma_i, c_i)$*,* $i = 1, \ldots, H$*, is the logistic function defined as*

$$f(s_t; \gamma_i, c_i) = \frac{1}{1 + e^{-\gamma_i (s_t - c_i)}}. \tag{2}$$

It is clear that $f(s_t; \gamma_i, c_i)$ is a monotonically increasing function such that $f(s_t; \gamma_i, c_i) \to 1$ as $s_t \to \infty$ and $f(s_t; \gamma_i, c_i) \to 0$ as $s_t \to -\infty$. The parameter γ_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*, which is assumed to be observable (measurable) at time $t - 1$. There are other possible choices for s_t; see Audrino and Trojani (2006), Chen, Chiang, and So (2003), and Scharth and Medeiros (2006) for some alternatives. To derive some of the results in Section 3 concerning stationarity, ergodicity, and the existence of moments of the FCGARCH model, and in the empirical applications, we consider the case where $s_t = y_{t-1}$. 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." However, the results concerning parameter estimation are derived without any particular choice of the transition variable.

The number of limiting regimes is defined by the hyperparameter H. For example, suppose that in (1), $s_t = y_{t-1}$, $H = 2$, c_1 is highly negative, and c_2

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

is very positive; then the resulting FCGARCH model will have three 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 $f(y_{t-1}; \gamma_i, c_i) \approx 0$, $i = 1, 2$. 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 $f(y_{t-1}; \gamma_1, c_1) \approx 1$ and $f(y_{t-1}; \gamma_2, c_2) \approx 0$. 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 $f(y_{t-1}; \gamma_i, c_i) \approx 1$, $i = 1, 2$. As the speed of the transitions between different limiting GARCH models is determined by the parameter γ_i, $i = 1, 2$, the multiple regime interpretation of the FCGARCH specification will become clearer the more abrupt are the transitions ($\gamma_i \gg 0$).[3] In practical applications, the restriction $\gamma_1 = \gamma_2 = \cdots = \gamma_H$ may be imposed 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 specifications that follow.

- 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 $s_t = h_{t-1}$
- The tree-structured GARCH model with a single threshold variable and no time-varying conditional mean if $\gamma_i \to \infty$, $i = 1, \ldots, H$

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), Carrasco and Chen (2002), Meitz (2005), and Francq and Zakoïan (2006) if $s_t = \varepsilon_{t-1}$. Finally, the FCGARCH model is encapsulated in the general nonlinear specification of Meitz and Saikkonen (2004).

3. MAIN ASSUMPTIONS AND PROBABILISTIC PROPERTIES OF THE FCGARCH MODEL

We need to make the following set of assumptions.

Assumption 1. The true and unique 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.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

Assumption 2. The sequence $\{\varepsilon_t\}$ of 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 zero.

Assumption 3. The parameters c_i and γ_i, $i = 1, \ldots, H$, satisfy the conditions

(R.1) $-\infty < c_1 < \cdots < c_H < \infty$;
(R.2) $\gamma_i > 0$.

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

Assumption 5. The parameters α_j, β_j, and λ_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 6. The process $\{s_t\}$ is strictly stationary and ergodic.

Assumption 1 is standard.[4] Assumption 2 is fundamental for the mathematical derivations in this section and in Section 5. Assumption 3 is important to guarantee 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. Assumption 6 is also related to the identification of the model parameters. If the transition variable is nonstationary some regimes may be not identifiable. Of course, Assumption 6 precludes time as a possible choice for s_t. For nonstationary transition variables the asymptotic theory should be modified in same spirit of Saikkonen and Choi (2004).

Set $f_{i,t} \equiv f(s_t; \gamma_i, c_i)$. Model (1) may be written as

$$\begin{aligned} y_t &= h_t^{1/2} \varepsilon_t, \\ h_t &= g_{t-1} + c_{t-1} h_{t-1}, \end{aligned} \tag{3}$$

where

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

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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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). Under Assumptions 2–5, the process $\mathbf{u}_t = (y_t, h_t)'$ is strictly stationary and ergodic if, and only if,*

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

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

$$\begin{aligned} y_t &= h_t^{1/2}\varepsilon_t, \\ h_t &= g_{t-1} + \sum_{k=0}^{\infty}\prod_{j=0}^{k} g_{t-1-k}c_{t-1-j}, \end{aligned} \tag{5}$$

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. Note also that we do not impose any particular choice for the transition variable.

COROLLARY 1. *Under the assumptions of Theorem 1 and $s_t = y_{t-1}$, $\forall t$, a sufficient condition for strict stationary and ergodicity of $\mathbf{u}_t = (y_t, h_t)'$ in terms of the parameters is*

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

Deriving a general sufficient condition for the existence of the moments of y_t is rather complicated. As mentioned in the previous section, the model families of He and Teräsvirta (1999), Ling and McAleer (2002), Lanne and Saikkonen (2005), Carrasco and Chen (2002), Meitz (2005), and Francq and Zakoïan (2006) do not nest the FCGARCH model without additional restrictions. Hence, the straightforward application of the results of these researchers is not possible. Although the FCGARCH model is nested in the general nonlinear model of Meitz and Saikkonen (2004), direct application of their results yields stricter conditions than the ones derived here; see, for example, the results in Table 1 in Meitz and Saikkonen (2004).

The following theorem states necessary conditions for the existence of moments of the FCGARCH model under no particular choice for the transition variable.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

TABLE 1. Simulated models: Descriptive statistics

	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

Note: The table shows descriptive statistics for the estimated kurtosis and first-order autocorrelation of the squared observations over 3,000 replications of models 1–3.

THEOREM 2. *Suppose that $y_t \in \mathbb{R}$ follows an FCGARCH(m,1,1) process as in (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*

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

In the subsequent corollary, we derive sufficient conditions for the existence of the second- and fourth-order moments of y_t in terms of the parameters and under the assumption that $s_t = y_{t-1}$.

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*

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

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

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

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.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

It is important to notice that even with explosive regimes the FCGARCH $(m,1,1)$ process 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 integrated GARCH (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 3,000 replications of the following FCGARCH(3,1,1) models with Gaussian errors, each of which has 5,000 observations.

Example 1

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

$$\begin{aligned} h_t &= 1\times 10^{-4} + 0.96h_{t-1} + 0.18y_{t-1}^2 \\ &\quad + \left(-0.9\times 10^{-4} - 0.60h_{t-1} - 0.10y_{t-1}^2\right) f\left(y_{t-1}; 5{,}000, -0.005\right) \\ &\quad + \left(1\times 10^{-4} + 0.10h_{t-1} + 0.05y_{t-1}^2\right) f\left(y_{t-1}; 5{,}000, 0.02\right). \end{aligned}$$

Example 2

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

$$\begin{aligned} h_t &= 6\times 10^{-5} + 1.10h_{t-1} + 0.10y_{t-1}^2 \\ &\quad + \left(-5\times 10^{-5} - 0.65h_{t-1} - 0.09y_{t-1}^2\right) f\left(y_{t-1}; 3{,}000, -0.005\right) \\ &\quad + \left(1\times 10^{-5} + 0.10h_{t-1} + 0.04y_{t-1}^2\right) f\left(y_{t-1}; 3{,}000, 0.005\right). \end{aligned}$$

Example 3

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

$$\begin{aligned} h_t &= 6\times 10^{-5} + 1.20h_{t-1} + 0.10y_{t-1}^2 \\ &\quad + \left(-5.5\times 10^{-5} - 1.20h_{t-1} - 0.10y_{t-1}^2\right) f\left(y_{t-1}; 2{,}000, -0.001\right) \\ &\quad + 5\times 10^{-5} f\left(y_{t-1}; 2{,}000, 0.01\right). \end{aligned}$$

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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 $\frac{1}{2}(\beta_0 + \lambda_0) + \frac{1}{2}\sum_{i=0}^{2}(\beta_i + \lambda_i) < 1$. 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 x's 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.

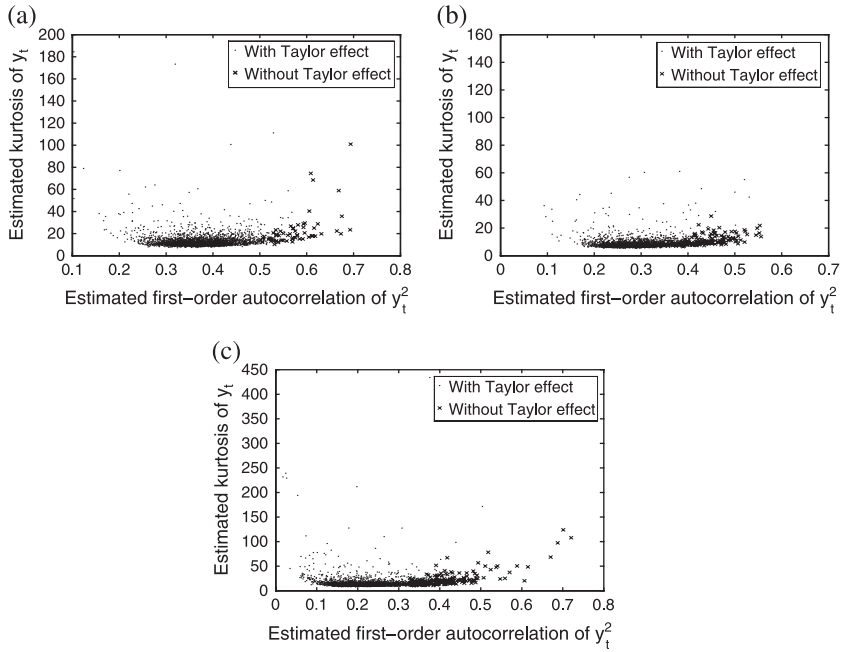


FIGURE 1. Scatter plot of the estimated kurtosis and first-order autocorrelation of the squared observations. The dots indicates the cases where the "Taylor effect" is satisfied, and the x's indicate the opposite effect. Panel (a) concerns Example 1. Panel (b) concerns Example 2. Panel (c) concerns Example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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

4. PARAMETER ESTIMATION

As the distribution of ε_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 vector autoregressive moving average–GARCH (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 (fourth-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. Finally, Straumann and Mikosch (2006) derived strong consistency and asymptotically normality of the QMLE of a general nonlinear GARCH(1,1) model using a stochastic recurrence equations approach.

As in Boussama (2000), McAleer, Chan, and Marinova (2007), Francq and Zakoïan (2004), and Straumann and Mikosch (2006) we prove consistency and asymptotic normality of the QMLE of the FCGARCH(m,1,1) under the log-moment condition in Theorem 1.[5] We also considered general transition variables. Applying the results in Straumann and Mikosch (2006) to the FCGARCH model is not possible unless we set the transition variable to be y_{t-1}, h_{t-1}, or ε_{t-1}. Finally, extending the results in Jensen and Rahbek (2004) for nonstationary 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

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

where $l_t(\boldsymbol{\psi}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_t) - (y_t^2/2h_t)$. 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 it is more convenient to work with the unobserved process $\left\{\left(y_{u,t}, h_{u,t}, s_{u,t}\right) : t = 0, \pm 1, \pm 2, \ldots\right\}$, which satisfies

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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

The unobserved quasi-log-likelihood function conditional on $\mathcal{F}_0 = \{(y_0, s_0), (y_{-1}, s_{-1}), \ldots\}$ is

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

with $l_{u,t}(\boldsymbol{\psi}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_{u,t}) - (y_{u,t}^2/2h_{u,t})$. 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 $\mathcal{L}(\boldsymbol{\psi}) = \mathsf{E}[l_{u,t}(\boldsymbol{\psi})]$. In Sections 4.1 and 4.2, we discuss the existence of $\mathcal{L}(\boldsymbol{\psi})$ and the identifiability of the FCGARCH model. Then, in Section 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 $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})\right| \xrightarrow{p} 0$, and the consistency of $\widehat{\boldsymbol{\psi}}_T$ follows. The asymptotic normality of both estimators is considered in Section 4.4. We start by proving asymptotic normality of $\widehat{\boldsymbol{\psi}}_{u,T}$. Then, using the results of Lemma 5 in Appendix B, 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. 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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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 "nonredundancy" in Hwang and Ding (1997); and the concept of reducibility of the model.

DEFINITION 2. *The FCGARCH*$(m,1,1)$ *model is minimal (or nonredundant) 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

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

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. *The FCGARCH model defined in (1) is reducible if one of the following three conditions holds.*

(i) *One of the triples* $(\alpha_i, \beta_i, \lambda_i)$ *vanishes jointly for any* $i \in [1, H]$*;*
(ii) $\gamma_i = 0$ *for some* $i \in [1, H]$*;*
(iii) *there is at least one pair* (i, j), $i \neq j$, $i = 1, \ldots, H$, $j = 1, \ldots, H$, *such that* $|\gamma_i (s_t - c_i)| = \left|\gamma_j \left(s_t - c_j\right)\right|$, $\forall s_t \in \mathbb{R}$, $t = 1, \ldots, T$ *(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 $f(s_t; \gamma_i, c_i) = 1 - f(s_t; -\gamma_i, c_i)$.
(P.3) Conditions (1) – (2) in the definition of reducibility provide information about the presence of irrelevant regimes, which translate into identifiability

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

sources. If the model contains a regime such that $\alpha_i = 0$, $\beta_i = 0$, and $\lambda_i = 0$, then the parameters γ_i and c_i remain unidentified for some $i \in [1, H]$. On the other hand, if $\gamma_i = 0$, then the parameters α_i, β_i, λ_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 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 ways 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 7. The parameters α_i, β_i, and λ_i do not vanish jointly for some $i \in [1, H]$.

Assumption 7 guarantees that there are no irrelevant regimes.

THEOREM 4. *Under Assumptions 3 and 7, 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 1–6,* $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{p} \boldsymbol{\psi}_0$ *and* $\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_0$.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

4.4. Asymptotic Normality

To prove asymptotic normality, we define

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

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

Consider the additional matrices

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

$$\mathbf{B}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\frac{\partial l_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'} = \frac{1}{4T}\sum_{t=1}^{T}\frac{1}{h_t^2}\left(\frac{y_t^2}{h_t}-1\right)^2\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'}. \tag{12}$$

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 1–6,*

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

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 that 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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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, even if convergence of the numerical optimizers is achieved.[6] Furthermore, 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 χ^2 distribution under the null.

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 we compute an upper bound to the overall significance level.[7]

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

Consider an FCGARCH with H limiting regimes, defined as

$$\begin{aligned} y_t &= h_t^{1/2}\varepsilon_t, \\ h_t &= \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \sum_{i=1}^{H-1}\left[\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2\right] f_{i,t}. \end{aligned} \tag{14}$$

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

$$\left[\alpha_H + \beta_H h_{t-1} + \lambda_H y_{t-1}^2\right] f_{H,t}. \tag{15}$$

A convenient null hypothesis is

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

against the alternative $\mathcal{H}_a : \gamma_H > 0$. Note that model (14) is not identified under the null hypothesis. To remedy this problem, we follow Luukkonen, Saikkonen, and Teräsvirta (1988) and expand the logistic function $f(s_t; \gamma_H, c_H)$ into a first-order Taylor expansion around the null hypothesis $\gamma_H = 0$.[8] After merging terms, the resulting model for h_t is

$$\begin{aligned} h_t = \tilde{\alpha}_0 + \tilde{\beta}_0 h_{t-1} + \tilde{\lambda}_0 y_{t-1}^2 + \sum_{i=1}^{H-1}\left[\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2\right] f_{i,t} \\ + \pi y_{t-1} + \delta h_{t-1} s_t + \rho y_{t-1}^2 s_t + R, \end{aligned} \tag{17}$$

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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

Under $\mathcal{H}_0$, $R = 0$, and the following result holds.

THEOREM 7. *Define* $\mathbf{v}_t = \left[y_{t-1}, h_{t-1}s_t, y_{t-1}^2 s_t\right]'$. *If the stationarity condition in Theorem 1 is satisfied, under Assumptions 2–6 and the additional assumption that* $\mathsf{E}\left[\mathbf{v}_t \mathbf{v}_t'\right] < \infty$ *under the null, the LM statistic given by*

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

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

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

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

$$\begin{aligned}\widehat{\mathbf{x}}_t = \Bigg[&1, \widehat{h}_{0,t-1}, y_{t-1}^2, f_{1,t}, \ldots, f_{H-1,t},\\ &f_{1,t}\widehat{h}_{0,t-1}, \ldots, f_{H-1,t}\widehat{h}_{0,t-1}, f_{1,t}y_{t-1}^2, \ldots, f_{H-1,t}y_{t-1}^2,\\ &(\widehat{\alpha}_1 + \widehat{\beta}_1 h_{0,t-1} + \widehat{\lambda}_1 y_{t-1}^2)\frac{\partial f_{1,t}}{\partial \gamma_1}, \ldots,\\ &(\widehat{\alpha}_{H-1} + \widehat{\beta}_{H-1}\widehat{h}_{0,t-1} + \widehat{\lambda}_{H-1}y_{t-1}^2)\frac{\partial f_{H-1,t}}{\partial \gamma_{H-1}},\\ &(\widehat{\alpha}_1 + \widehat{\beta}_1 h_{0,t-1} + \widehat{\lambda}_1 y_{t-1}^2)\frac{\partial f_{1,t}}{\partial c_1}, \ldots,\\ &(\widehat{\alpha}_{H-1} + \widehat{\beta}_{H-1}h_{0,t-1} + \widehat{\lambda}_{H-1}y_{t-1}^2)\frac{f_{H-1,t}}{\partial c_{H-1}}\Bigg]',\end{aligned}$$

$$\widehat{\mathbf{v}}_t = \left[y_{t-1}, \widehat{h}_{0,t-1}s_t, y_{t-1}^2 s_t\right]',$$

and

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

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

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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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 $SSR_0 = \sum_{t=1}^{T} \left(y_t^2 / \widehat{h}_{0,t} - 1 \right)^2$.
2. Regress $\left(y_t^2 / \widehat{h}_{0,t} - 1 \right)$ 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

$$LM = T \frac{SSR_0 - SSR_1}{SSR_0}, \tag{19}$$

or the F statistic

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

Under $\mathcal{H}_0$, LM is asymptotically distributed as χ^2 with 3 degrees of freedom, and F has an asymptotic 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 before.
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 $\left(\left(y_t^2 / \widehat{h}_{0,t} \right) - 1 \right) \widehat{\mathbf{r}}_t$ and compute the residual sum of squares, SSR. The test statistic given by

$$LM_R = T - SSR \tag{21}$$

has an asymptotic χ^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, it is 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).

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

6. MODELING CYCLE

We are now ready to combine the preceding statistical ingredients into a practical modeling strategy. We begin by testing linearity against an ARCH(q) model at significance level δ using the Engle (1982) LM test.[9] The model under the null hypothesis is a homoskedastic model. If the null hypothesis is not rejected, the homoskedastic model is considered as the data generating process (DGP). In case of rejection, a GARCH(1,1) model is estimated and tested against an FCGARCH 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 nonrejection 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 ϱ is ad hoc. To avoid selecting small models (few regimes), it is good practice to carry-out the modeling cycle with different values of ϱ. In the empirical examples, we consider $\varrho = 1/2$ and $\varrho = 1/3$. The results are the same in both cases.

The estimation of the FCGARCH model is computationally demanding, especially the selection of starting values. For this paper, we use a method based on a genetic algorithm to estimate the initial values for the parameters. With this careful selection of initial values, convergence is usually achieved.

7. MONTE-CARLO EXPERIMENT

The purpose of this section is to check the performance of the test described in Section 5. We use four DGPs, considering $s_t = y_{t-1}$. The DGPs are as follows.

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): $\alpha_0 = 1 \times 10^{-4}$, $\beta_0 = 0.96$, $\lambda_0 = 0.18$, $\alpha_1 = -0.9 \times 10^{-4}$, $\beta_1 = -0.60$, $\lambda_1 = -0.10$, $\alpha_2 = 1 \times 10^{-4}$, $\beta_2 = 0.10$, $\lambda_2 = 0.05$, $\gamma_1 = 5{,}000$, $\gamma_2 = 5{,}000$, $c_1 = -0.005$, and $c_2 = 0.02$.
4. Model D
   FCGARCH(3,1,1): $\alpha_0 = 6 \times 10^{-5}$, $\beta_0 = 1.10$, $\lambda_0 = 0.10$, $\alpha_1 = -5 \times 10^{-5}$, $\beta_1 = -0.65$, $\lambda_1 = -0.09$, $\alpha_2 = 1 \times 10^{-5}$, $\beta_2 = 0.10$, $\lambda_2 = 0.04$, $\gamma_1 = 300$, $\gamma_2 = 6{,}000$, $c_1 = -0.005$, and $c_2 = 0.025$.

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.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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, whereas model B does not.

TABLE 2. Simulation: Modeling strategy results

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
		≥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
		≥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
		≥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
		≥4	0	0	0	0	0.004
C	Gaussian	1	0	0.024	0	0	0
		2	0.952	0.976	0.936	0.868	0.772
		3	0.048	0	0.064	0.116	0.192
		≥4	0	0	0	0.016	0.036
	t with 10 d.f.	1	0	0.044	0.012	0	0
		2	0.884	0.956	0.912	0.888	0.848
		3	0.108	0	0.072	0.104	0.140
		≥4	0.008	0	0.004	0.008	0.012
D	Gaussian	1	0.084	0.104	0	0	0
		2	0.912	0.896	0.872	0.800	0.716
		3	0.004	0	0.128	0.196	0.268
		≥4	0	0	0	0.004	0.016
	t with 10 d.f.	1	0.108	0.140	0	0.000	0
		2	0.812	0.860	0.880	0.808	0.748
		3	0.080	0	0.120	0.192	0.240
		≥4	0	0	0	0	0.012

Note: The table reports the frequency with which a model with a given number of limiting regimes is selected over 1,000 simulations. A thousand observations of each model are simulated at each replication. In all the simulations the parameter ϱ equals 2. The value δ is the initial significance level of the sequence of LM tests. The abbreviation d.f. = degrees of freedom.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

We include model B in our simulation to evaluate the effect of the violation of the assumption that $\mathrm{E}[\mathbf{v}_t\mathbf{v}_t'] < \infty$ in the behavior of the test. Models C and D are different specifications of an FCGARCH(3,1,1) model with y_{t-1} as transition variable. Model C was 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 condition $\mathrm{E}[\mathbf{v}_t\mathbf{v}_t'] < \infty$. Model C has two sharp transitions with, on average, 10% and 1% of the observations along the first and second transitions, respectively. On the other hand, model D has one transition that is very smooth and a second one that is sharp. On average, 90% of the observations lie along the first transition, whereas only 0.3% of the observations are along the second transition. To evaluate small-sample properties of the test, all the simulations are based on series with 1,000 observations.[10] To avoid any initialization effect we generate data with 1,500 observations and discard the first 500 observations; see Lundbergh and Teräsvirta (2002). A total of 1,000 replications have been generated. Only the results concerning the robust version of the tests are shown 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 Schwarz's Bayesian information criterion (SBIC). The sequence of LM tests is carried out with three different initial significance levels α. The value of the hyperparameter ϱ 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, especially 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).

8. EMPIRICAL EXAMPLES

We consider daily logarithm returns of 10 stock indexes: AEX (The Netherlands), ATX (Austria), CAC40 (France), DAX (Germany), FTSE100 (United Kingdom), Hang Seng (Hong Kong), IBOVESPA (Brazil), Nikkei (Japan), SMI (Switzerland), and S&P500 (United States). These indexes are chosen 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. Table 3 shows, for each series, the sample period and also some descriptive statistics. Here μ is the sample mean, σ is the sample standard deviation, SK is the sample skewness, and K is the sample kurtosis.

For each series we estimate the standard GARCH(1,1) model and also the GJR(1,1) specification. The estimation is performed using the Bollerslev–Wooldridge QML approach and the Marquardt algorithm. The adequacy of these models is then checked using the sign bias, negative size bias, positive size bias,

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

TABLE 3. Daily returns: Descriptive statistics

	In-sample			Out-of-sample			Statistics (in-sample)			
Series	Start	End	Obs.	Start	End	Obs.	μ	σ	SK	K
AEX	10/12/91	12/31/02	2,590	01/02/03	01/08/07	1,031	0.00	0.01	−0.24	7.12
ATX	11/11/91	12/30/02	2,501	01/02/03	01/08/07	996	0.00	0.01	−0.63	7.54
CAC40	03/01/90	12/31/02	3,218	01/02/03	01/08/07	1,031	0.00	0.01	−0.12	5.46
DAX	11/26/90	12/30/02	3,037	01/02/03	01/08/07	1,027	0.00	0.01	−0.24	6.63
FTSE100	01/02/90	12/31/02	3,284	01/02/03	01/08/07	1,016	0.00	0.01	−0.09	5.58
Hang Seng	01/02/90	12/31/02	3,217	01/02/03	01/08/07	999	0.00	0.02	−0.02	11.92
IBOVESPA	04/27/93	12/30/02	2,393	01/02/03	01/08/07	996	0.00	0.03	0.53	10.92
Nikkei	01/04/90	12/30/02	3,203	01/06/03	01/08/07	986	0.00	0.02	0.26	6.12
SMI	11/09/90	12/30/02	3,050	01/03/03	01/08/07	1,014	0.00	0.01	−0.23	7.71
S&P500	01/02/90	12/31/02	3,280	01/02/03	01/08/07	1,011	0.00	0.01	−0.12	6.81

Note: The table shows the sample period, the number of observations, and descriptive statistics for the daily returns. Here μ is the sample mean, σ is the sample standard deviation, SK is the sample skewness, and K is the sample kurtosis.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

TABLE 4. *Estimated models and specification tests*

	GARCH(1,1) model: $h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 r_{t-1}^2$														
Series	α_0	β_0	λ_0	σ	SK	K	Sb	Nsb	Psb	Jsb	Pc	Pi	Pcc	AIC	Likelihood
AEX	1.24×10^{-6} (3.18×10^{-7})	0.90 (0.01)	0.09 (0.01)	1.00	−0.33	3.77	0.00	0.07	0.00	0.00	0.89	0.79	0.96	−6.21	8.05×10^3
ATX	3.71×10^{-6} (1.17×10^{-6})	0.88 (0.03)	0.08 (0.02)	1.00	−0.39	4.98	0.03	0.00	0.03	0.00	0.79	0.86	0.95	−6.44	8.06×10^3
CAC40	5.31×10^{-6} (2.22×10^{-6})	0.89 (0.02)	0.08 (0.01)	1.00	−0.32	4.74	0.09	0.14	0.01	0.05	0.58	0.54	0.71	−5.89	9.48×10^3
DAX	3.90×10^{-6} (2.07×10^{-6})	0.90 (0.01)	0.08 (0.01)	1.00	−0.77	11.32	0.02	0.46	0.02	0.06	0.13	0.67	0.30	−5.94	9.02×10^3
FTSE100	1.46×10^{-6} (3.56×10^{-7})	0.91 (0.01)	0.08 (0.01)	1.00	−0.09	3.80	0.65	0.56	0.11	0.38	0.41	0.93	0.71	−6.50	1.07×10^4
Hang Seng	7.12×10^{-6} (2.78×10^{-6})	0.88 (0.02)	0.10 (0.02)	1.00	−0.53	7.08	0.03	0.00	0.02	0.00	0.68	0.40	0.64	−5.56	8.94×10^3
IBOVESPA	1.77×10^{-5} (4.26×10^{-6})	0.84 (0.02)	0.14 (0.02)	1.00	−0.25	4.44	0.00	0.00	0.00	0.00	0.27	0.95	0.54	−4.57	5.47×10^3
Nikkei	7.23×10^{-6} (2.20×10^{-6})	0.87 (0.02)	0.10 (0.01)	1.00	0.03	4.79	0.05	0.08	0.01	0.04	0.43	0.32	0.44	−5.64	9.03×10^3
SMI	7.31×10^{-6} (4.00×10^{-6})	0.81 (0.04)	0.13 (0.02)	1.00	−0.73	9.83	0.00	0.01	0.01	0.01	0.54	0.88	0.82	−6.31	9.63×10^3
S&P500	5.98×10^{-7} (1.54×10^{-7})	0.93 (0.01)	0.06 (0.01)	1.00	−0.39	5.11	0.00	0.00	0.00	0.00	0.21	0.97	0.45	−6.54	1.07×10^4

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

TABLE 4. *Continued.*

	GJR-GARCH(1,1) model: $h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 r_{t-1}^2 + \lambda_1 r_{t-1}^2 I(r_{t-1} \le 0)$																
Series	α_0	β_0	λ_0	λ_1	σ	SK	K	Sb	Nsb	Psb	Jsb	Pc	Pi	Pcc	AIC	Likelihood
AEX	1.59×10^{-6} (3.78×10^{-7})	0.90 (0.03)	0.05 (0.02)	0.07 (0.03)	1.00	−0.33	3.98	0.03	0.53	0.00	0.01	0.62	0.40	0.62	−6.22	8.06×10^3
ATX	3.99×10^{-6} (9.84×10^{-7})	0.89 (0.02)	0.04 (0.01)	0.08 (0.02)	1.00	−0.28	4.68	0.21	0.03	0.31	0.20	0.93	0.79	0.96	−6.45	8.07×10^3
CAC40	4.73×10^{-6} (1.35×10^{-6})	0.92 (0.01)	0.01 (0.01)	0.09 (0.02)	1.00	−0.33	5.03	0.54	0.74	0.17	0.57	0.63	0.35	0.58	−5.91	9.51×10^3
DAX	4.23×10^{-6} (1.50×10^{-6})	0.90 (0.01)	0.03 (0.02)	0.08 (0.02)	1.00	−0.74	11.23	0.08	0.91	0.06	0.13	0.06	0.80	0.17	−5.95	9.03×10^3
FTSE100	1.26×10^{-6} (2.55×10^{-7})	0.93 (0.01)	0.02 (0.01)	0.08 (0.01)	1.00	−0.08	3.96	0.66	0.57	0.85	0.88	0.14	0.18	0.13	−6.51	1.07×10^4
Hang Seng	8.20×10^{-6} (2.44×10^{-6})	0.88 (0.02)	0.03 (0.02)	0.11 (0.02)	1.00	−0.43	6.65	0.25	0.15	0.23	0.45	0.42	0.11	0.20	−5.58	8.98×10^3
IBOVESPA	1.85×10^{-5} (4.25×10^{-6})	0.84 (0.02)	0.10 (0.02)	0.10 (0.03)	1.00	−0.18	4.34	0.01	0.00	0.01	0.01	0.53	0.47	0.63	−4.58	5.48×10^3
Nikkei	5.57×10^{-6} (1.60×10^{-6})	0.89 (0.01)	0.02 (0.01)	0.13 (0.02)	1.00	0.04	4.73	0.54	0.69	0.22	0.48	0.93	0.73	0.94	−5.65	9.07×10^3
SMI	7.48×10^{-6} (2.64×10^{-6})	0.82 (0.02)	0.03 (0.02)	0.18 (0.04)	1.00	−0.60	8.80	0.06	0.36	0.12	0.28	0.59	0.19	0.38	−6.34	9.67×10^3
S&P500	1.34×10^{-6} (2.69×10^{-7})	0.92 (0.01)	0.01 (0.01)	0.12 (0.02)	1.00	−0.39	5.11	0.00	0.19	0.00	0.00	0.38	0.22	0.32	−6.57	1.08×10^4

Note: The table reports the estimation and diagnostic test results of GARCH and GJR models for the daily returns. The estimation is performed by the QML. The numbers in parentheses below the estimates are the Bollerslev–Wooldridge robust standard errors. The terms σ, SK, and K are the standard deviation, the skewness, and the kurtosis of the standardized residuals, respectively. The terms 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). The terms Pc, Pi, and Pcc are the p-values of the tests of unconditional coverage, independence, and conditional 95% coverage. AIC is the value of Akaike's information criteriom and Likelihood is the estimated log-likelihood.

and joint tests for asymmetry proposed by Engle and Ng (1993). Table 4 reports the estimation and diagnostic test results of GARCH and GJR models for the daily returns. The numbers in parentheses below the estimates are the Bollerslev–Wooldridge robust standard errors. The terms *Sb*, *Nsb*, *Psb*, and *Jsb* are the p-values of the sign bias, negative size bias, positive size bias, and joint tests for asymmetry, *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 4, the normalized residuals from GJR specification have equivalent kurtosis to the ones from the GARCH alternative. On the other hand, the skewness coefficients are somewhat closer to zero for the GJR model. The values of the AIC for all estimated GJR models are lower than the ones for the GARCH models. Other interesting facts emerge from the table. First, the sum of the estimated β_0 and λ_0 coefficients in the GARCH 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 size bias, positive size bias, and joint tests it is clear that there are asymmetric effects in the normalized residuals from the GARCH models. The analysis of the coverage tests indicates that the GARCH models seem to produce correct confidence intervals.

When the GJR model is considered, it is important to mention that a negative shock induces an explosive regime, as the sum of the estimated β_0, λ_0, and λ_1 parameters is greater than one; see Table 6. The parameter λ_1 is significant for all series. Concerning the results of the sign bias, negative size bias, positive size bias, and joint tests it seems that there are still some asymmetric effects (at the 10% level) in five out of the 10 series considered, namely, AEX, DAX, IBOVESPA, SMI, and S&P500. Finally, apart from the DAX series, the results of the coverage tests indicate that the GJR models do provide correct interval estimation.

We proceed with 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. At each step of the testing sequence we halve the significance level of the test ($\varrho = 1/2$). We also carryout the test sequence with other values for ϱ, and the results do not change. The initial significance level for the sequence of LM tests is 5%. Table 5 shows the estimation results and diagnostic statistics. The estimation is performed by the QML method using the sequential quadratic programming numerical optimization algorithm. To avoid convergence problems, we divide the transition variable, y_{t-1}, by its unconditional standard deviation. The numbers in parentheses below the estimates are the standard errors.

The sequence of robust LM tests shows evidence of two limiting regimes for six series: AEX, ATX, FTSE100, Hang Seng, Nikkei, and SMI. For the CAC40, DAX, IBOVESPA, and S&P500 three limiting regimes are found. It is important to mention that for the AEX and Nikkei series the parameter c_1 is positive and

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

statistically different from zero. On the other hand, for the ATX, Hang Seng, and SMI indexes c_1 is negative and statistically significant. Both cases contradict the usual zero threshold considered in the literature. Only for the FTSE100 series is the value of c_1 not statistically different from zero. For the series with three regimes it is clear that the first limiting (extreme) regime is associated with very negative shocks, representing "very bad news." The middle regime is related to tranquil periods, and the third and extreme regime represents large positive shocks. The slope parameter is not very high for nine series, indicating a smooth transition. Only for the Hang Seng index, the estimated transition is abrupt. In this case, we do not report the standard errors for the slope parameter because they are not very accurate. Moreover, as pointed out in Section 5, the t-statistic does not have its customary distribution under the null hypothesis that $\gamma = 0$.

Comparing the AIC from the FCGARCH model with the one from the GARCH and GJR specifications, the FCGARCH outperforms the other two alternatives, indicating that the final model is not overparametrized. Observing the results in Table 5 it is clear that, for most of the series, the estimated standardized residuals from the FCGARCH model have kurtosis coefficients slightly lower than both the GARCH and GJR models. As in the GJR case, the FCGARCH model seems to describe adequately the asymmetric relation between returns and volatility. In addition, when the coverage tests are considered, the FCGARCH model produces correct confidence intervals for all series with the sole exception of the DAX index.

One interesting fact is the large value of the estimated β_0's, which indicates a very persistent regime associated with negative returns. Table 6 shows the persistence associated with each limiting regime in both the GJR and FCGARCH models. Considering the GJR 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 on whether 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 "good news regime" depending on whether the estimated model has two or three regimes. Finally the last column in the table shows the persistence of the last limiting regime in the FCGARCH model and is associated with "very good news." 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 specification. Second, the GARCH effect seems to be dissipated when the returns become more positive, especially 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 FCGARCH models. We use both the mean absolute errors (MAEs) and the root mean squared errors (RMSEs) as performance measures. The squared returns are used as a proxy to the volatility.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

TABLE 5. FCGARCH estimation and specification tests

	Estimation results												
Series	α_0	α_1	α_2	β_0	β_1	β_2	λ_0	λ_1	λ_2	γ_1	γ_2	c_1	c_2
AEX	2.22×10^{-16} (2.50×10^{-3})	2.36×10^{-6} (3.35×10^{-7})	—	1.26 (0.05)	−0.80 (0.01)	—	0.11 (0.06)	−0.05 (0.05)	—	0.47 (0.14)	—	0.54 (0.08)	—
ATX	9.08×10^{-6} (6.85×10^{-9})	-1.69×10^{-6} (1.25×10^{-7})	—	1.06 (0.03)	−0.25 (0.04)	—	0.07 (0.03)	−0.03 (0.01)	—	4.04 (1.67)	—	−0.79 (0.07)	—
CAC40	2.22×10^{-16} (1.20×10^{-4})	-3.94×10^{-33} (1.73×10^{-5})	2.60×10^{-5} (8.31×10^{-4})	1.26 (0.07)	−0.28 (0.04)	−0.33 (0.07)	0.05 (0.04)	−0.05 (0.01)	0.04 (0.01)	3.50 (1.65)	3.01 (1.51)	−1.27 (0.03)	0.87 (0.03)
DAX	1.27×10^{-16} (1.25×10^{-6})	-2.34×10^{-6} (4.24×10^{-7})	2.76×10^{-5} (1.23×10^{-5})	1.22 (0.10)	−0.30 (0.03)	−0.21 (0.06)	0.04 (0.01)	−0.04 (0.01)	0.02 (0.01)	5.00 (1.24)	7.97 (1.98)	−0.81 (0.08)	0.72 (0.10)
FTSE100	1.27×10^{-4} (1.92×10^{-5})	-1.65×10^{-6} (2.79×10^{-5})	—	1.22 (0.17)	−0.56 (0.12)	—	0.07 (0.02)	−0.03 (0.03)	—	0.56 (0.14)	—	−0.05 (0.04)	—
Hang Seng	2.22×10^{-16} (2.92×10^{-6})	-1.25×10^{-5} (3.88×10^{-6})	—	0.99 (0.30)	−0.15 (0.04)	—	0.11 (0.04)	−0.08 (0.04)	—	300 (−−)	—	−0.31 (0.07)	—
IBOVESPA	2.22×10^{-16} (1.88×10^{-6})	-1.25×10^{-5} (3.92×10^{-6})	1.00×10^{-3} (9.81×10^{-4})	1.49 (0.05)	−0.60 (0.01)	−0.73 (0.08)	0.13 (0.04)	−0.04 (0.01)	−0.05 (0.01)	9.95 (6.21)	7.70 (3.45)	−1.82 (0.01)	2.48 (0.02)
Nikkei	7.37×10^{-6} (3.71×10^{-6})	-7.36×10^{-6} (2.43×10^{-6})	—	1.16 (0.13)	−0.53 (0.12)	—	0.08 (0.02)	−0.02 (0.01)	—	1.11 (0.13)	—	0.17 (0.01)	—
SMI	2.22×10^{-16} (1.18×10^{-6})	-1.37×10^{-5} (1.41×10^{-6})	—	1.15 (0.36)	−0.47 (0.31)	—	0.11 (0.02)	−0.07 (0.02)	—	4.00 (1.15)	—	−0.40 (0.06)	—
S&P500	9.77×10^{-16} (3.72×10^{-14})	-5.14×10^{-7} (1.51×10^{-7})	1.81×10^{-5} (1.22×10^{-8})	1.21 (0.11)	−0.32 (0.03)	−0.25 (0.08)	0.06 (0.02)	−0.01 (0.07)	−0.04 (0.02)	2.52 (1.01)	2.85 (1.23)	−0.72 (0.02)	1.56 (0.09)

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

TABLE 5. *Continued.*

	Diagnostic test results											
Series	σ	SK	K	Sb	Nsb	Psb	Jsb	Pc	Pi	Pcc	AIC	Likelihood
AEX	1.00	−0.32	3.95	0.21	0.67	0.02	0.03	0.89	0.79	0.96	−6.23	8.07×10^3
ATX	1.00	−0.24	4.69	0.62	0.13	0.87	0.44	0.52	0.69	0.75	−6.45	8.07×10^3
CAC40	1.00	−0.29	4.60	0.97	0.76	0.50	0.81	0.47	0.49	0.61	−5.92	9.53×10^3
DAX	1.00	−0.71	10.8	0.18	0.70	0.09	0.14	0.01	0.41	0.01	−5.96	9.07×10^3
FTSE100	1.00	−0.09	3.96	0.31	0.17	0.41	0.59	0.32	0.46	0.47	−6.52	1.07×10^4
Hang Seng	1.00	−0.41	6.67	0.41	0.40	0.39	0.78	0.56	0.10	0.19	−5.58	8.97×10^3
IBOVESPA	1.00	−0.25	4.25	0.00	0.01	0.00	0.01	0.97	0.23	0.49	−4.58	5.50×10^3
Nikkei	1.00	0.06	4.65	0.63	0.27	0.46	0.36	0.80	0.50	0.78	−5.67	9.08×10^3
SMI	1.00	−0.57	8.36	0.42	0.66	0.35	0.81	0.84	0.54	0.81	−6.35	9.70×10^3
S&P500	1.00	−0.33	4.67	0.01	0.24	0.00	0.01	0.75	0.32	0.58	−6.58	1.08×10^4

Note: The table reports the estimation and diagnostic test results of the FCGARCH model for the unpredictable daily returns. The estimation is performed by the QML method. The numbers in parentheses below the estimates are the standard errors. The terms σ, 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). The term 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: $h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \left(\alpha_1 + \beta_1 h_{t-1} + \lambda_1 y_{t-1}^2\right) f\left(r_{t-1}/\sigma; \gamma_1, c_1\right) + \left(\alpha_2 + \beta_2 h_{t-1} + \lambda_2 y_{t-1}^2\right) f\left(y_{t-1}/\sigma; \gamma_2, c_2\right)$, where σ is the unconditional standard deviation of y_{t-1}.

TABLE 6. GJR and FCGARCH models: Persistence in each regime

	GJR model		FCGARCH model		
Series	"Bad news" $\beta_0+\lambda_0+\lambda_1$	"Good news" $\beta_0+\lambda_0$	$\beta_0+\lambda_0$	$\beta_0+\beta_1+\lambda_0+\lambda_1$	$\beta_0+\beta_1+\beta_2+\lambda_0+\lambda_1+\lambda_2$
AEX	1.02	0.95	1.38	0.53	—
ATX	1.01	0.93	1.14	0.86	—
CAC40	1.02	0.93	1.31	0.98	0.69
DAX	1.01	0.93	1.26	0.92	0.73
FTSE100	1.03	0.95	1.30	0.70	—
Hang Seng	1.02	0.91	1.11	0.88	—
IBOVESPA	1.04	0.94	1.62	0.98	0.20
Nikkei	1.04	0.91	1.23	0.69	—
SMI	1.09	0.91	1.30	0.73	—
S&P500	1.05	0.93	1.27	0.94	0.64

Note: The table shows the persistence associated with each limiting regime in both the GJR and FCGARCH models. The sum $\beta_0+\lambda_0+\lambda_1$ is the persistence associated with negative past returns in the GJR model ("bad news"), whereas $\beta_0+\lambda_0$ 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 on whether 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 "good news regime" depending on whether the estimated model has two or three regimes. Finally, the last column in the table shows the persistence of the last limiting regime in the FCGARCH model with three regimes, which is associated with "very good news."

The results are shown in Table 7. We compare the performance of the FCGARCH model with the GARCH, GJR, EGARCH, and tree-structured GARCH models. Analyzing the results, on average, the FCGARCH model outperforms the concurrent specifications. More specifically, the FCGARCH model attains the lowest MAE for four series and the lowest RMSEs in seven cases. In two cases, the FCGARCH and the EGARCH are the best models according to the MAE figures and in one case according to the RMSE. The EGARCH, tree-structured GARCH, and FCGARCH models are even for one series when the MAE is considered, and the FCGARCH and tree-structured GARCH specifications are equivalent in two cases when the RMSE is used.

To check whether the forecasts are statistically different we use the superior predictive ability (SPA) test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any other competing models in terms of a given loss function. When the MAE is used we have the following results. The GARCH is the only model rejected by the SPA test in the case of the AEX series. The forecast performance of the FCGARCH model is statistically inferior to concurrent specifications in only two cases: the Hang Seng and S&P500 indexes. For the first one, the only model that produces adequate forecasts is the EGARCH model. For the S&P500 series, the GARCH, GJR, and tree-structured

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

TABLE 7. Forecasting performance

	MAE				
Series	GARCH	GJR	EGARCH	TSGARCH	FCGARCH
AEX	1.56 (0.04)	1.53 (0.20)	1.52 (0.36)	1.53 (0.18)	1.49 (0.63)
ATX	1.04 (0.00)	1.03 (0.03)	1.00 (0.71)	1.03 (0.01)	1.00 (0.59)
CAC40	1.28 (0.00)	1.27 (0.00)	1.21 (0.86)	1.29 (0.00)	1.22 (0.41)
DAX	1.67 (0.00)	1.64 (0.02)	1.62 (0.87)	1.62 (0.76)	1.62 (0.19)
FTSE100	0.74 (0.00)	0.73 (0.00)	0.72 (0.70)	0.72 (0.03)	0.71 (0.72)
Hang Seng	1.21 (0.00)	1.22 (0.00)	1.16 (0.70)	1.19 (0.00)	1.19 (0.00)
IBOVESPA	2.99 (0.00)	2.99 (0.00)	3.03 (0.00)	2.95 (0.01)	2.88 (0.77)
Nikkei	1.57 (0.00)	1.53 (0.00)	1.49 (0.00)	1.50 (0.00)	1.44 (0.77)
SMI	1.00 (0.03)	0.99 (0.05)	0.97 (0.81)	1.00 (0.00)	0.97 (0.77)
S&P500	0.65 (0.13)	0.64 (0.92)	0.66 (0.00)	0.65 (0.14)	0.66 (0.04)
	RMSE				
Series	GARCH	GJR	EGARCH	TSGARCH	FCGARCH
AEX	4.11 (0.13)	4.03 (0.51)	4.03 (0.48)	4.05 (0.20)	3.83 (0.60)
ATX	2.63 (0.21)	2.60 (0.32)	2.60 (0.47)	2.61 (0.13)	2.58 (0.81)
CAC40	2.69 (0.27)	2.70 (0.06)	2.67 (0.53)	2.69 (0.18)	2.58 (0.63)
DAX	3.45 (0.16)	3.40 (0.96)	3.41 (0.55)	3.40 (0.20)	3.39 (0.55)
FTSE100	1.69 (0.09)	1.65 (0.15)	1.66 (0.24)	1.66 (0.22)	1.61 (0.63)
Hang Seng	1.68 (0.05)	1.68 (0.00)	1.67 (0.00)	1.66 (0.53)	1.66 (0.25)
IBOVESPA	4.16 (0.00)	4.16 (0.08)	4.20 (0.00)	4.10 (0.73)	4.10 (0.57)
Nikkei	2.42 (0.00)	2.42 (0.01)	2.39 (0.12)	2.40 (0.02)	2.38 (0.78)
SMI	2.13 (0.46)	2.12 (0.46)	2.11 (0.76)	2.12 (0.66)	2.10 (0.74)
S&P500	1.09 (0.62)	1.09 (0.90)	1.09 (0.88)	1.10 (0.07)	1.09 (0.30)

Note: The table shows the mean absolute errors (MAE) and the root mean squared errors (RMSE) 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. "TSGARCH" denotes the tree-structured GARCH model. The numbers in parentheses are the p-values of the superior predictive ability (SPA) test developed by Hansen (2005).

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

GARCH have similar performances. On the other hand, for the ATX, CAC40, FTSE100, and SMI series, the FCGARCH and EGARCH specifications produce superior forecasts. For the DAX series, the EGARCH, tree-structured GARCH, and FCGARCH models have similar performance. Finally, for the IBOVESPA and the Nikkei, the FCGARCH alternative is the only model that produces adequate forecasts.

When the RMSE is considered, most of the models produce similar forecasts according to the SPA test. However, some points deserve mention. The FCGARCH model is the only one that does not have the forecasts rejected for any of the 10 series. On the other hand, the forecasts of the tree-structured GARCH model are statistically inferior in two cases: the Nikkei and S&P500 indexes. When the Hang Seng and IBOVESPA indexes are considered, the only models that are not rejected are the tree-structured GARCH and FCGARCH models.

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 and also intermittent dynamics. The model is called the flexible coefficient GARCH (FCGARCH) and is a 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 were 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 QMLE mator were derived under second- and fourth-order moment conditions. A modeling cycle based on a sequence of simple and easily implemented LM tests was discussed to avoid the estimation of unidentified models. A Monte-Carlo experiment was designed to evaluate the methodology, and it was shown that the modeling strategy works well in moderate samples.

An empirical example with 10 stock indexes showed evidence of two limiting regimes for six series and three limiting regimes for the other four. In addition, for all the series with three limiting regimes, the first limiting (extreme) regime was associated with very negative shocks, representing "very bad news." 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 was extremely explosive, indicating that bad news may induce very high volatility. When a forecasting exercise was considered, the FCGARCH outperformed the GARCH, GJR, EGARCH, and tree-structured GARCH alternatives.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

NOTES

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

2. In addition, as shown in the empirical examples, first-order models seem to be adequate to describe the time dependence in the data considered in this paper.

3. 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).

4. Note that in Assumption 1 we assume that H is known. Hence, testing for the number of regimes is, in principle, precluded. Therefore, Assumption 1 will be dropped when deriving the results in Section 5.

5. See also Li, Ling, and McAleer (2002) and McAleer (2005).

6. 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.

7. An equivalent procedure has been adopted in Medeiros and Veiga (2005) and Medeiros, Teräsvirta, and Rech (2006).

8. Lundbergh and Teräsvirta (2002) applied the same ideas to evaluate GARCH models.

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

10. A thousand observations is a rather small sample in the conditional volatility literature.

REFERENCES

Anderson, H.M., K. Nam, & F. Vahid (1999) Asymmetric nonlinear smooth transition GARCH models. In P. Rothman (ed.) *Nonlinear Time Series Analysis of Economic and Financial Data*, pp. 191–207. Kluwer.

Audrino, F. & P. Bühlmann (2001) Tree-structured GARCH models. *Journal of the Royal Statistical Society, Series B* 63, 727–744.

Audrino, F. & F. Trojani (2006) Estimating and predicting multivariate volatility regimes in global stock markets. *Journal of Applied Econometrics* 21, 345–369.

Bacon, D.W. & D.G. Watts (1971) Estimating the transition between two intersecting lines. *Biometrika* 58, 525–534.

Berkes, I., L. Horváth, & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–227.

Black, F. (1976) Studies in stock price volatility changes. *ASA Proceedings, Business and Economic Statistics*, 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 d'un modéle 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 & Economic Statistics* 12, 309–316.

Carnero, M.A., D. Peña, & E. Ruiz (2004) Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics* 2, 319–342.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chan, K.S. & H. Tong (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 179–190.

Chen, C.W.S., T.C. Chiang, & M.K.P. 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.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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

Comte, F. & O. Lieberman (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.

Ding, Z., C.W.J. Granger, & R.F. Engle (1993) A long memory property 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 inflation. *Econometrica* 50, 987–1007.

Engle, R.F. & V.K. Ng (1993) Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749–1778.

Fornari, F. & 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. & J.-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. & J.-M. Zakoïan (2006) Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed processes. *Econometric Theory* 22, 815–834.

Glosten, L., R. Jagannanthan, & 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*, 2nd ed. Econometric Society Monographs 16. Cambridge University Press.

Goetzmann, W.N., R.G. Ibbotson, & 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, 61–78.

Gourieroux, C. & A. Monfort (1995) *Statistics and Econometric Models*, vol. 2. Cambridge University Press.

Granger, C.W.J. & 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. Ph.D. Dissertation, Stockholm School of Economics.

Hamilton, J.D. & 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.

Hansen, P.R. (2005) A test for superior predictive ability. *Journal of Business & Economic Statistics* 23, 365–380.

He, C. & T. Teräsvirta (1999) Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92, 173–192.

Hwang, J.T.G. & 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. & A. Rahbek (2004) Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* 72, 641–646.

Lanne, M. & P. Saikkonen (2005) Nonlinear GARCH models for highly persistent volatility. *Econometrics Journal* 8, 251–276.

Lee, S.-W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Li, C.W. & 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, & M. McAleer (2002) Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* 16, 245–269.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

Ling, S. & M. McAleer (2002) Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* 106, 109–117.
Ling, S. & M. McAleer (2003) Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.
Liu, J., W.K. Li & C.W. Li (1997) On a threshold autoregression with conditional heteroscedastic variances. *Journal of Statistical Planning and Inference* 62, 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. & T. Teräsvirta (2002) Evaluating GARCH models. *Journal of Econometrics* 110, 417–435.
Luukkonen, R., P. Saikkonen, & T. Teräsvirta (1988) Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–499.
Malmsten, H. & 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, & D. Marinova (2007) An econometric analysis of asymmetric volatility: Theory and application to patents. *Journal of Econometrics* 139, 259–284.
Medeiros, M.C., T. Teräsvirta, & G. Rech (2006) Building neural network models for time series: A statistical approach. *Journal of Forecasting* 25, 49–75.
Medeiros, M.C. & A. Veiga (2005) A flexible coefficient smooth transition time series model. *IEEE Transactions on Neural Networks* 16, 97–113.
Meitz, M. (2005) A necessary and sufficient condition for the strict stationarity of a family of GARCH processes. *Econometric Theory* 22, 985–988.
Meitz, M. & P. Saikkonen (2004) Ergodicity, Mixing, and Existence of Moments of a Class of Markov Models with Applications to GARCH and ACD Models. Working Paper Series in Economics and Finance 573, Stockholm School of Economics.
Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.
Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
Newey, W. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. Elsevier Science.
Rabemananjara, R. & J.M. Zakoian (1993) Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics* 8, 31–49.
Saikkonen, P. & I. Choi (2004) Cointegrating smooth transition regressions. *Econometric Theory* 20, 301–340.
Scharth, M. & M.C. Medeiros (2006) Asymmetric Effects and Long Memory in the Volatility of Dow Jones Stocks. Discussion paper 532, Pontifical Catholic University of Rio de Janeiro.
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.
Straumann, D. & T. Mikosch (2006) Quasi-MLE in heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34, 2449–2495.
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. & I. Mellin (1986) Model selection criteria and model selection tests in regression models. *Scandinavian Journal of Statistics* 13, 159–171.
Trapletti, A., F. Leisch, & K. Hornik (2000) Stationary and integrated autoregressive neural network processes. *Neural Computation* 12, 2427–2450.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

van Dijk, D. & P.H. Franses (1999) Modelling multiple regimes in the business cycle. *Macroeconomic Dynamics* 3, 311–340.
van Dijk, D., P.H. Franses, & A. Lucas (1999a) Testing for ARCH in the presence of additive outliers. *Journal of Business & Economic Statistics* 17, 217–235.
van Dijk, D., P.H. Franses, & A. Lucas (1999b) Testing for smooth transition nonlinearity in the presence of outliers. *Journal of Applied Econometrics* 14, 539–562.
White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press.
White, H. (2001) *Asymptotic Theory for Econometricians*. Academic Press.
Wooldridge, J.M. (1990) A unified approach to robust, regression-based specification tests. *Econometric Theory* 6, 17–43.
Wooldridge, J.M. (1994) Estimation and inference for dependent process. In R.F. Engle & D.L. McFadden (eds.) *Handbook of Econometrics*, vol. 4, pp. 2639–2738. Elsevier Science.
Zakoïan, J.M. (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.

APPENDIX A: Proofs of Theorems and Corollaries

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

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

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,

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

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 δ 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 α-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, Cor. 3.48, p. 49). Hence, the remainder of the proof is identical to the one of Theorem 2 in Nelson (1990). This completes the proof. ■

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

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

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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

Note that, because

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

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

Take a large value for the constants M and N and write the following expected valuer:

$$\begin{aligned}\mathsf{E}\left[\log(c_{t-1})\right]&=\mathsf{E}\left[\log(c_{t-1})\left|I_{|\varepsilon_{t-1}|<\left(M/h_{t-1}^{1/2}\right)}\right.\right]\mathsf{Pr}\left[I_{|\varepsilon_{t-1}|<\left(M/h_{t-1}^{1/2}\right)}\right]\\&\quad+\mathsf{E}\left[\log(c_{t-1})\left|I_{|\varepsilon_{t-1}|\ge\left(M/h_{t-1}^{1/2}\right)}\right.\right]\mathsf{Pr}\left[I_{|\varepsilon_{t-1}|\ge\left(M/h_{t-1}^{1/2}\right)}\right]\\&=\mathsf{E}\left[\log(c_{t-1})\left|I_{|\varepsilon_{t-1}|<\left(M/h_{t-1}^{1/2}\right)}\right.\right]p\\&\quad+\mathsf{E}\left[\log(c_{t-1})\left|I_{|\varepsilon_{t-1}|\ge\left(M/h_{t-1}^{1/2}\right)}\right.\right](1-p).\end{aligned}$$

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

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

The log-moment condition can be expressed as

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

However,

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

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

$$\mathsf{E}\left[(c_{t-1})\left|I_{|\varepsilon_{t-1}|\ge\left(M/h_{t-1}^{1/2}\right)}\right.\right]\le 1.$$

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

Then, as δ_λ and δ_β can be made arbitrarily small,

$$\begin{aligned}
&\mathsf{E}\left[(c_{t-1})\middle| I_{|\varepsilon_{t-1}|\geq\left(M/h_{t-1}^{1/2}\right)}\right] \\
&\quad=\frac{\mathsf{E}\left[\beta_U+\varepsilon_{t-1}^2\lambda_U|I_{|\varepsilon_{t-1}|\geq\left(M/h_{t-1}^{1/2}\right)}\right]+\mathsf{E}\left[\beta_0+\varepsilon_{t-1}^2\lambda_0\middle| I_{|\varepsilon_{t-1}|<\left(M/h_{t-1}^{1/2}\right)}\right]}{2} \\
&\quad\leq\frac{\beta_U+\lambda_U+\beta_0+\lambda_0}{2}.
\end{aligned}$$

Then, if

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

the log-moment condition holds. ■

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

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

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

$$\mathbf{u}_t=\mathbf{a}_t+\mathbf{C}_{t-1}\mathbf{u}_{t-1}, \tag{A.4}$$

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

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

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

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

because $a_{t,j}\leq\left(\alpha_0+\sum_{i=1}^{H}|\alpha_i|\right)^j<\infty$. Here $\mathcal{F}_{t-2}$ 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

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

■

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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

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

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 to save space.

This completes the proof. ■

Proof of Theorem 3. It is easy to see that $G(\mathbf{w}_t;\boldsymbol{\psi})$ 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 γ_i and c_i. Similarly, we can see that $G(\mathbf{w}_t;\boldsymbol{\psi})$ 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 $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|h_{u,t}\right|\right]<\infty$. By Jensen's inequality $\mathsf{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\ln\left|h_{u,t}\right|\right|\right]<\infty$. Thus $\mathsf{E}\left[\left|l_{u,t}(\boldsymbol{\psi})\right|\right]<\infty$, $\forall\boldsymbol{\psi}\in\boldsymbol{\Psi}$. This completes the proof. ■

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

$$\boldsymbol{\phi}_0'\mathbf{z}_t+\sum_{i=1}^{H}\boldsymbol{\phi}_i'\mathbf{z}_t f(s_t;\boldsymbol{\theta}_i)=\widetilde{\boldsymbol{\phi}}_0'\mathbf{z}_t+\sum_{i=1}^{H}\widetilde{\boldsymbol{\phi}}_i'\mathbf{z}_t f(s_t;\widetilde{\boldsymbol{\theta}}_i). \tag{A.8}$$

To show global identifiability of the FCGARCH model, we need to prove that, under restrictions (R.1) and (R.2) in Assumptions 3 and if Assumption 7 holds, (A.8) is satisfied if, and only if, $\boldsymbol{\psi}=\widetilde{\boldsymbol{\psi}}$.

Equation (A.8) can be rewritten as

$$\boldsymbol{\phi}_0'\mathbf{z}_t-\widetilde{\boldsymbol{\phi}}_0'\mathbf{z}_t-\sum_{j=1}^{2H}\overline{\boldsymbol{\phi}}_j'\mathbf{z}_t f\left(s_t;\overline{\boldsymbol{\theta}}_j\right)=0, \tag{A.9}$$

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 $j=H+1,\ldots,2H$.

Let $\varphi_j=\gamma_j\left(s_t-c_j\right)$, $j=1,\ldots,2H$. Lemma 2.7 in Hwang and Ding (1997) implies that if φ_{j_1} and φ_{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 7 precludes that possibility. Hence, φ_{j_1} and φ_{j_2} must be sign-equivalent. But restriction (R.2) in Assumption 4 precludes that two functions φ_{j_1} and φ_{j_2} coming from the same model will be sign-equivalent. Consequently, $\exists\, j_1\in\{0,\ldots,H\}$ and $j_2\in\{H+1,\ldots,2H+1\}$ such

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

that φ_{j_1} and φ_{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 $\boldsymbol{\phi}_i = \widetilde{\boldsymbol{\phi}}_i$ 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

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

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. ■

Proof of Theorem 5. Following Newey and McFadden (1994) (See also White (1994) and Wooldridge (1994)), $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{p} \boldsymbol{\psi}_0$ if the following conditions hold.

(a) The parameter space $\boldsymbol{\Psi}$ is compact.

(b) $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ 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 $\boldsymbol{\psi} \in \boldsymbol{\Psi}$.

(c) $\mathcal{L}(\boldsymbol{\psi})$ has a unique maximum at $\boldsymbol{\psi}_0$.

(d) $\mathcal{L}_{u,T}(\boldsymbol{\psi}) \xrightarrow{p} \mathcal{L}(\boldsymbol{\psi})$.

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

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

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

Proof of Theorem 6. We start by proving asymptotical 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 asymptotical normality of the QMLE we need the following conditions in addition to the ones stated in the proof of Theorem 5.

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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

(b) 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 $\boldsymbol{\Psi}$.

(c) The matrix $\mathbf{A}_T(\boldsymbol{\psi}) \xrightarrow{p} \mathbf{A}(\boldsymbol{\psi}_0)$ for any sequence $\boldsymbol{\psi}_T$ such that $\boldsymbol{\psi}_T \xrightarrow{p} \boldsymbol{\psi}_0$.

(d) The score vector satisfies

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

Condition (a) is satisfied by assumption. Condition (b) 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 implies that condition (c) is satisfied. Furthermore, nonsingularity of $\mathbf{A}(\boldsymbol{\psi}_0)$ follows immediately from identification of the FCGARCH model and the nonsingularity of $\mathbf{B}(\boldsymbol{\psi}_0)$; see Hwang and Ding (1997). In Lemma 4 in Appendix B we prove that condition (d) is also met. This completes the proof. ■

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

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

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

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

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

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

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

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

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

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

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

see Engle (1982).

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

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

$$LM = \mathbf{q}(\boldsymbol{\psi})|_{\mathcal{H}_0} \left[\mathbf{A}_T(\boldsymbol{\psi})|_{\mathcal{H}_0}\right]^{-1} \mathbf{q}(\boldsymbol{\psi})|_{\mathcal{H}_0}$$
$$= \frac{T}{2} \sum_{t=1}^{T} \left(\frac{y_t^2}{h_t} - 1\right) \mathbf{d}_t \left[\sum_{t=1}^{T} \mathbf{d}_t \mathbf{d}_t'\right]^{-1} \sum_{t=1}^{T} \left(\frac{y_t^2}{h_t} - 1\right) \mathbf{d}_t. \quad \textbf{(A.13)}$$

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

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*

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

Proof. The proof follows the same reasoning as the one from Lemma 5 in Lumsdaine (1996). Define $\boldsymbol{\xi}_t = \frac{\partial h_t}{\partial \boldsymbol{\psi}}$. It is straightforward to show that

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

where

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

and $\overline{\beta}(s_t) = \beta_0 + \sum_{i=1}^{H} \beta_i f_{i,t}$. Then, $\mathbf{d}'\boldsymbol{\xi}_t = \mathbf{d}'\overline{\beta}(s_t)\boldsymbol{\xi}_{t-1} + \mathbf{d}'\boldsymbol{\kappa}_{t-1}$. Because 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. Because $\boldsymbol{\kappa}_t$ is nondegenerate, $\mathbf{d}'\boldsymbol{\xi}_t = 0$ with probability one if and only if $\mathbf{d} = \mathbf{0}$. This completes the proof. ■

LEMMA 2. *Under the assumptions of Lemma 1,*

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

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 $\mathbf{Y}_t = [y_t, y_{t-1}, y_{t-2}, \ldots, s_t, s_{t-1}, s_{t-2}, \ldots]'$. Theorem 3 implies that $\mathsf{E}\left[\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |g(\mathbf{Y}_t, \boldsymbol{\psi})|\right] < \infty$. In addition, because

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

$g(\mathbf{Y}_t, \boldsymbol{\psi})$ is stationary with $\mathsf{E}[g(\mathbf{Y}_t, \boldsymbol{\psi})] = 0$, 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. ■

LEMMA 3. *Under the assumptions of Lemma 1*

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

Proof. The proof of this lemma is adapted from the proof of item (i) of Lemma 6 in Lumsdaine (1996). It is clear that under the condition of Theorem 1,

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

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

Set $\overline{\beta}(s_t) = \beta_0 + \sum_{i=1}^{H} \beta_i f_{i,t}$. Combining (9) and (11) we get

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

Therefore, dividing the left hand side of the preceding expression by h_t, we have

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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

Define $X_t \equiv y_t^2$ and $\xi_t \equiv \left| \left(\prod_{j=1}^{t} \overline{\beta}(s_j) \right) (h_0 - h_{u,0}) \right|$. Under the condition of Theorem 1 X_t is a strictly stationary and ergodic time series, with $\mathsf{E}[|X_t|] < \infty$. Furthermore, it is clear by Theorem 1 that $\sup_t |\xi_t| \leq C$, 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,

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

This completes the proof. ■

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

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

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

$$\nabla_0 l_{u,t} \equiv \frac{\partial l_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \Big|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \quad \text{and} \quad \nabla_0 h_{u,t} \equiv \frac{\partial h_{u,t}}{\partial \boldsymbol{\psi}} \Big|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}. \tag{B.1}$$

Then,

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

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

$$\mathsf{E}\left[\nabla_0 l_{u,t} \nabla_0 l_{u,t}' \right] \leq \frac{1}{4} \mathbf{M} \mathsf{E}\left[\left(\varepsilon_t^4 - 2\varepsilon_t^2 + 1 \right) \right] = \frac{1}{4} \mathbf{M} (\mu_4 - 1), \tag{B.3}$$

where $\mu_4 = \mathsf{E}\left[\varepsilon_t^4 \right] < \infty$. Hence, $\mathsf{E}\left[\nabla_0 l_{u,t} \nabla_0 l_{u,t}' \right] \leq \infty$ 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 and s_t. By the given assumptions $\mathsf{E}[S_T] > 0$. Using the central limit theorem of Stout (1974),

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

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press

Finally, by the Cramér–Wold device,

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

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. ■

LEMMA 5. *Under the conditions of Theorem 6*

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

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

Proof. First set

$$\nabla_0^2 l_{u,t} \equiv \left.\frac{\partial^2 l_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \quad \text{and} \quad \nabla_0^2 h_{u,t} \equiv \left.\frac{\partial^2 h_{u,t}}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}. \tag{B.4}$$

Then,

$$\nabla_0^2 l_{u,t} = \left(\frac{y_t^2}{h_{u,t}} - 1\right)\frac{1}{2h_{u,t}}\nabla_0^2 h_{u,t} - \frac{1}{2h_{u,t}^2}\left(\frac{y_t^2}{h_{u,t}}\right)\nabla_0 h_{u,t}\nabla_0 h_{u,t}'. \tag{B.5}$$

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

$$\nabla_0^2 l_{u,t} \le \left(\varepsilon_t^2 - 1\right)\frac{1}{\underline{\delta}}\mathbf{M}_2 - \frac{1}{2\underline{\delta}^2}\varepsilon_t^2\mathbf{M} \tag{B.6}$$

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

This completes the proof. ■

https://doi.org/10.1017/S026646660809004X Published online by Cambridge University Press