Studies in Nonlinear Dynamics & Econometrics

Volume 12, *Issue* 3 2008 *Article* 3

REGIME-SWITCHING MODELS IN ECONOMICS AND FINANCE

Optimal Test for Markov Switching GARCH Models

Liang Hu* Yongcheol Shin†

*University of Leeds, l.hu@leeds.ac.uk
†University of Leeds, ys@lubs.leeds.ac.uk

Optimal Test for Markov Switching GARCH Models*

Liang Hu and Yongcheol Shin

Abstract

Empirically, the sum of GARCH parameter estimates is found to be close to unity, suggesting that the conditional volatility of most stock return data are likely to follow an integrated GARCH (IGARCH) process. However, such an extremely high persistence in unconditional variance may be overstated because of neglected structural breaks or parameter changes. As a result it is important to distinguish between these two processes, one being a globally stationary process and the other being a nonstationary IGARCH process. Though there are a number of studies modelling asymmetry leverage effects and advancing a battery of specification tests, studies that directly propose specification tests against Markov switching (MS) GARCH models are almost nonexistent. This paper develops such tests against MS-GARCH processes, which is shown to be asymptotically equivalent to the LR test. Furthermore, we consider the case in which the conditional variance follows an IGARCH process under the null whilst it is globally stationary under the alternative. Monte Carlo studies show that our proposed tests have a good finite sample performance. In an application to the weekly stock return data for five East Asian emerging markets, we find strong evidence in favor of MS-GARCH models.

*We are grateful to the editor, two anonymous referees, Charlie Cai, Marine Carrasco, Kausik Chaudhuri, George Kapetanios, Changjin Kim, Inmoo Kim, Taewhan Kim, Werner Ploberger, Kevin Reilly, Robert Sollis, Sam Yu and seminar participants at the Korean Economic Association International Conference at Seoul National University, the ESRC Seminar Series by Nonlinear Economics and Finance Research Community at Keele University, the 2007 Econometric Society European Meeting at Budapest, the University of Leeds, Yonsei University, Queen Mary College at the University of London and the University of Newcastle for their helpful comments. The usual disclaimer applies.

1. Introduction

Empirically, a routine finding is that the sum of GARCH parameter estimates is close to unity, indicating that most stock return data follow an integrated GARCH (IGARCH) process and thus the associated unconditional variance will be likely to be unbounded. Lamoureux and Lastrapes (1990) argue that the extremely high persistence in the unconditional variance may be overstated because of neglected structural breaks. Caporale, Pittis and Spagnolo (2003) show via Mote Carlo studies that fitting misspecified GARCH models to the data generated by a (stationary) Markov Switching (MS) GARCH process tends to produce IGARCH parameter estimates. Analytically, Maekawa *et al.* (2004) show that the structural change in the form of Markov Switching parameter changes in GARCH models can mislead analysts to choose an IGARCH model. See also Francq and Zakoïan (2001), Mikosch and Starica (2004) and Hillebrand (2005).

As a result it is important to distinguish between these two processes, one being a regime-dependent stationary process and the other being a nonstationary IGARCH process.[1] Just as in the case of the standard unit root tests, the problem would be trivial asymptotically as one process is stationary whereas the other is not. As is well known from the unit root test literature (*e.g.* Perron, 1989), however, linear nonstationary and nonlinear stationary processes may be confused in small samples to such an extent that the standard tests based on the linear models are unable to distinguish between them. Despite such a growing concern, very little research has been devoted to tests for structural break or parameter constancy in GARCH models, especially against IGARCH models. See Lee *et al.* (2003) and Smith (2006).

The main interest of this paper lies in modelling and testing MS-GARCH models. MS-GARCH models explain reasonably well the stylized facts that conditional volatility can increase substantially in a short period of turbulent times and there exists asymmetric leverage effects of shocks on conditional volatility by estimating high and low volatility regimes, respectively. For the growing number of the MS-GARCH family see Hamilton and Susmel (1994), Cai (1994), Gray (1996), Dueker (1997), Garcia (1998), Francq *et al.* (2001) and Haas *et al.* (2004).

A class of nonlinear asymmetric models has already been proposed in the literature and assumed great significance, e.g., GJR-GARCH by Glosten, Jagannathan and Runkle (1993), Threshold GARCH of Zakoian (1994), Quadratic

[1] As shown in Nelson (1990), IGARCH is not covariance stationary, but it is strictly stationary. In this paper, by stationarity we mean covariance stationary.

GARCH by Sentana (1995) and STAR-GARCH by Chan and McAleer (2002). There is also a number of studies proposing specification tests for (stationary) GARCH models against a variety of asymmetric models stated above, e.g. Engle and Ng (1993), Hagerud (1997), Lundeberg and Teräsvirta (2002) and Haulunga and Orme (2005).

However, studies that directly address the specification tests against MS-GARCH models are almost nonexistent. This surprising absence is due to the fact that estimation and inference for MS-GARCH models will become very complicated due to their highly nonlinear path-dependent structure. Recently, Carrasco, Hu and Ploberger (2005, CHP thereafter) propose a general testing procedure for parameter constancy in the MS framework. The test shares the optimal property of the information matrix-based tests. Importantly, we only need to estimate the model under the null, which facilitates the computation of the test statistic. This is a great advantage since estimating MS-GARCH model is computationally intractable due to high nonlinearity and path dependence. This paper follows CHP and proposes an optimal testing procedure to distinguish linear GARCH/IGARCH from a class of MS-GARCH models. Namely, with the null hypothesis of GARCH, the test could serve as a misspecification test for possible nonlinearity and regime dependency in the data. While with the null hypothesis of IGARCH, our test would be optimal in testing whether it is the regime-switching mechanism that leads to an overstated unconditional variance, thus a false conclusion of IGARCH. In a broad context, we also argue that our proposed test has an optimal power in detecting whether it is the structural break or parameter changes in GARCH models that may lead to overstatement of conditional volatility parameters.

Chaudhuri and Klaassen (2001) use a regime switching GARCH model for stock returns of the five East Asian countries and show that such a specification gives better results because it allows for both volatility persistence and the possibility of the structural change in the conditional variance occurring as a result of the East Asian crisis. Not surprisingly, they do not provide any specification tests against MS-GARCH models. In an application to the weekly stock returns data for five East Asian emerging markets we find strong evidence in favor of MS-GARCH models and thus provide a support for their empirical findings. Considering that the sum of the single-regime GARCH parameter estimates is all close to unity and the robust Wald test is not able to reject the null hypothesis of IGARCH, we may conclude that neglecting the parameter changes would result in the spurious finding of IGARCH estimates, a finding consistent with Caporale, Pittis and Spagnolo (2003) and Hillebrand (2005).

The plan of the paper is as follows: Section 2 discusses the models and the assumptions. Section 3 proposes the optimal testing procedure and discusses the asymptotic distribution property. Section 4 evaluates the small sample performance of the proposed tests and demonstrates that their small sample power is reasonably satisfactory. Misspecification issues are also investigated. Section 5 presents an empirical application. Section 6 concludes. Mathematical proofs are collected in an Appendix.

2. Model and Assumptions

We begin with the standard GARCH$(1,1)$ regression model:

$$\begin{cases} y_t = \mu_t + v_t \\ v_t = \sigma_t \varepsilon_t,\ \varepsilon_t \sim iid(0,1) \\ \sigma_t^2 = \alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \end{cases} \tag{2.1}$$

where μ_t is the time-varying conditional mean possibly given by $\boldsymbol{\beta}'\mathbf{x}_t$ with $\mathbf{x}_t$ being the $k \times 1$ vector of stochastic covariates. We are interested in testing the null of GARCH$(1,1)$ model, (2.1) against the alternative of an MS-GARCH$(1,1)$ model given by

$$\sigma_t^2 = \alpha_{0,S_t} + \alpha_{1,S_t} v_{t-1}^2 + \gamma_{1,S_t} \sigma_{t-1}^2, \tag{2.2}$$

where we allow the conditional variance, σ_t^2, to be subject to a hidden Markov chain, S_t.

There are relatively few studies that directly address the specification tests against MS-GARCH models. There are two main difficulties. First, the nuisance parameters specifying the MS mechanism are not identified under the null, which makes the testing problem nonstandard. Secondly, estimation of the MS-GARCH model will be computationally intractable due to the high nonlinearity and the path-dependent structure; namely, σ_t^2 in (2.2) depends on the entire history of regimes. There have been a few studies to circumvent the path dependence. For example, Gray (1996) suggests that the conditional variance at time $t-1$ given information at time $t-2$, can be integrated out and then the likelihood function can be evaluated using a first-order recursive scheme as in the same way as in the basic regime switching model of Hamilton (1989). See also Dueker (1997) and Haas *et al.* (2004) for alternative estimation algorithms.

In order to circumvent the difficulties described above, we follow the optimal testing procedure for parameter constancy in the MS framework proposed by CHP. We define the parameter vector of interest, $\boldsymbol{\theta} = (\mu, \alpha_0, \alpha_1, \gamma_1)'$, and denote the vector of parameter changes under the alternative by

$$\boldsymbol{\eta}_t = c\boldsymbol{\delta} S_t, \tag{2.3}$$

where S_t is a scalar Markov chain, $\boldsymbol{\delta}$ is a vector specifying the direction of the alternative and c is a scalar specifying the amplitude of the change. Moreover, we have $Corr\left(S_t, S_s\right) = \rho^{|t-s|}$ for some $|\rho| < 1$. This specification includes a wide range of models. For example, assuming that there are two states, denoted $S_t = \{0, 1\}$, we consider a case where the parameters in (2.2) are given by $\alpha_{00} = \alpha_0$, $\alpha_{01} = \alpha_0 + \upsilon_1$; $\alpha_{10} = \alpha_1$, $\alpha_{11} = \alpha_1 + \upsilon_2$; and $\gamma_{10} = \gamma_1$, $\gamma_{11} = \gamma_1 + \upsilon_3$. That is, under the alternative, one regime is the same as the null while the other regime takes a different specification. Then we can express the parameter changes as in (2.3), where

$$c = \sqrt{\upsilon_1^2 + \upsilon_2^2 + \upsilon_3^2};\ \boldsymbol{\delta} = \left(\frac{\upsilon_1}{\sqrt{\upsilon_1^2 + \upsilon_2^2 + \upsilon_3^2}}, \frac{\upsilon_2}{\sqrt{\upsilon_1^2 + \upsilon_2^2 + \upsilon_3^2}}, \frac{\upsilon_3}{\sqrt{\upsilon_1^2 + \upsilon_2^2 + \upsilon_3^2}}\right)'$$

We define the nuisance parameter vector, $\boldsymbol{\lambda} = (c, \boldsymbol{\delta}', \rho)'$, where $\boldsymbol{\delta}$ is normalized by $\|\boldsymbol{\delta}\| = 1$ for identification, and make the following assumptions:

Assumption 1. S_t is a scalar geometric ergodic Markov chain with an n-dimensional state space and $Var\left(S_t\right) = 1$.

Assumption 2. $\boldsymbol{\lambda}$ belongs to a compact set, Λ.

Assumption 3. The innovation ε_t is standard normal and is independent of the state variable S_t.

Assumption 1 and 2 are from CHP. Assumption 3 is commonly used in GARCH models.[2] We also require the following conditions to hold:[3]

$$\begin{aligned} \sup_{t,\theta\in\mathcal{N}} E\left(\left\|\ell_t^{(1)}\right\|^{24}\right) &< \infty;\ \sup_{t,\theta\in\mathcal{N}} E\left(\left\|\ell_t^{(2)}\right\|^{12}\right) < \infty;\ \sup_{t,\theta\in\mathcal{N}} E\left(\left\|\ell_t^{(3)}\right\|^{8}\right) < \infty; \\ \sup_{t,\theta\in\mathcal{N}} E\left(\left\|\ell_t^{(4)}\right\|\right) &< \infty;\ \sup_{t,\theta\in\mathcal{N}} E\left(\left\|\ell_t^{(5)}\right\|\right) < \infty, \end{aligned}$$

where $\ell_t = \ell_t\left(\boldsymbol{\theta}\right)$ denotes the conditional log-density of y_t given $y_{t-1}, ..., y_1$ under the null, $\ell_t^{(r)} = \ell_t^{(r)}\left(\boldsymbol{\theta}\right)$ is its r-th derivative with respect to $\boldsymbol{\theta}$ and $\mathcal{N}$ is a neighborhood around the true parameter vector, $\boldsymbol{\theta}_0$. Thus, $\ell_t\left(\boldsymbol{\theta}\right)$ should be at least five times differentiable. Notice under Assumption 3 that these conditions are trivially satisfied.

[2]In Section 4 we consider the case where ε_t follows the Student's t-distribution and investigate the impact of the fat-tailed distribution by stochastic simulations.

[3]This is a sufficient condition for establishing that our proposed test will be asymptotically equivalent to the LR test. In certain situations these moment conditions can be weakened. See CHP.

3. Optimal Test Statistic

Using the (modified) Information Matrix-based testing principle proposed by CHP, we obtain the optimal test statistic for the null of a linear GARCH model, (2.1) against the alternative of MS-GARCH$(1,1)$ model, (2.2) by

$$\Psi_T(\boldsymbol{\lambda}) = \Psi_T\left(\boldsymbol{\lambda};\hat{\boldsymbol{\theta}}\right) = \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\xi_t\left(\boldsymbol{\lambda};\hat{\boldsymbol{\theta}}\right) - \frac{1}{2T}\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\lambda})'\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\lambda}), \tag{3.1}$$

where

$$\xi_t(\boldsymbol{\lambda};\boldsymbol{\theta}) = c^2\boldsymbol{\delta}'\left[\left(\frac{\partial^2 l_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \left(\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)'\right) + 2\sum_{s=1}^{t-1}\rho^{(t-s)}\left(\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial l_s}{\partial\boldsymbol{\theta}}\right)'\right]\boldsymbol{\delta}, \tag{3.2}$$

$\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\lambda})$ is the residual from the regression of $\frac{1}{2}\xi_t\left(\boldsymbol{\lambda};\hat{\boldsymbol{\theta}}\right)$ on $\ell_t^{(1)}\left(\hat{\boldsymbol{\theta}}\right)$, and $\hat{\boldsymbol{\theta}}$ is the quasi-maximum likelihood estimator of $\boldsymbol{\theta}$ obtained under the null of GARCH(1,1) model. Notice that the test statistic, $\Psi_T(\boldsymbol{\lambda})$ depends only on the QMLE of $\boldsymbol{\theta}$ obtained under the null.

When nuisance parameters, $\boldsymbol{\lambda} = (c, \boldsymbol{\delta}', \rho)'$, are unknown, the test procedure will suffer from the Davies (1987) problem since $\boldsymbol{\lambda}$'s are not identified under the null. Most solutions to this problem involve integrating out unidentified parameters from the test statistics. This is usually achieved by constructing the summary statistics over a grid set of $\boldsymbol{\lambda}$. Following Andrews and Ploberger (1994) and Hansen (1996) we consider the exponential average and the supremum of the statistic defined respectively by

$$\Psi_{\exp} = \int_{\bar{\Lambda}} \exp\left(\Psi_T(\boldsymbol{\lambda})\right) dJ(\boldsymbol{\lambda}) \text{ and } \Psi_{\sup} = \sup_{\boldsymbol{\lambda}\in\bar{\Lambda}} \Psi_T(\boldsymbol{\lambda}), \tag{3.3}$$

where $J(\boldsymbol{\lambda})$ is some prior distribution for $\boldsymbol{\lambda}$ with support $\bar{\Lambda}$ on a compact subset of $\Lambda = \{c^2, \boldsymbol{\delta}, \rho : c^2 > 0, \|\boldsymbol{\delta}\| = 1, \underline{\rho} < \rho < \bar{\rho}\}$ with $-1 < \underline{\rho} < \bar{\rho} < 1$.[4]

Theorem 3.1. *Let Assumptions 1-3 hold. Then,* $\Psi_{\exp}$ *and* $\Psi_{\sup}$*statistics defined in (3.3) are admissible.*

A general proof of Theorem 3.1 follows directly from CHP. Here we thus provide a brief discussion on this asymptotic result. Our test statistics share

[4] To avoid the case where the test statistic grows without bound, we need to rule out the case that ρ is on the boundary of the parameter space.

similar form to the Information Matrix test. To see this, the second-order Bartlett identity leads to an information matrix equality in (3.2),

$$E\left(\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}+\left(\frac{\partial l_t}{\partial\theta_i}\right)\left(\frac{\partial l_t}{\partial\theta_j}\right)\right)=0,\ \boldsymbol{\theta}\in\mathcal{R}^p,\ i,j=1,...,p.$$

White (1982) proposes the information matrix test of the following form:

$$T\mathbf{D}_T'Var\left(\mathbf{D}_T\right)^{-1}\mathbf{D}_T\xrightarrow{d}\chi_p^2,$$

where $\mathbf{D}_T=T^{-1}\sum_{t=1}^T\mathbf{d}_t$ and $\mathbf{d}_t$ is the vector with elements $\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}+\left(\frac{\partial l_t}{\partial\theta_i}\right)\left(\frac{\partial l_t}{\partial\theta_j}\right)$. Chesher (1984) advances the information matrix test for neglected heterogeneity, which is shown to be equivalent to an LM test for the null hypothesis that the variance of the random coefficient is zero. It is nonstandard since the parameter falls on the boundary of the parameter space under the null where variance is equal to zero. It is shown that the score with respect to this variance becomes

$$\sum_{t=1}^T c^2\boldsymbol{\delta}'\left(\frac{\partial^2 l_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}+\left(\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)'\right)\boldsymbol{\delta},\tag{3.4}$$

which is very similar to $\xi_t(\boldsymbol{\lambda};\boldsymbol{\theta})$ defined in (3.2) above.

Our test is admissible in the sense that it is asymptotically equivalent to the likelihood-ratio test statistic.[5] Since the nuisance parameters, $\boldsymbol{\lambda}$ are not identified under the null hypothesis, there exists no uniformly most powerful test in this case. So we now need to show that our test is asymptotically equivalent to the Likelihood Ratio test for a specific sequence of alternatives, namely,

$$H_{1T}:\boldsymbol{\theta}_t=\boldsymbol{\theta}_0+T^{-1/4}\boldsymbol{\eta}_t-T^{-1/2}\mathbf{d}\left(\boldsymbol{\lambda},\boldsymbol{\theta}_0\right),$$

where $\mathbf{d}(\boldsymbol{\lambda},\boldsymbol{\theta}_0)=(I(\boldsymbol{\theta}_0))^{-1}\,cov\left(\frac{1}{2}\xi_t(\boldsymbol{\lambda};\boldsymbol{\theta}_0),l_t^{(1)}(\boldsymbol{\theta}_0)\right)$, and $I(\boldsymbol{\theta}_0)$ is the information matrix evaluated at the (unknown) true parameters, $\boldsymbol{\theta}_0$.

The main difference between our proposed test given by (3.1)-(3.2) and the information matrix type tests based on (3.4), is that our test has an extra term, $2\sum_{s=1}^{t-1}\rho^{(t-s)}\left(\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)'$. This is due to the fact that we do not assume that the parameters are independent but allow them to be serially correlated. Notice that this term becomes a Martingale difference sequence, when evaluated at the QML estimate of $\boldsymbol{\theta}$ under the null. Thus, our test shares the same property of Information Matrix-based tests. Moreover, the term $\frac{1}{2T}\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\lambda})'\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\lambda})$ is an

[5]Our proposed test can be of a Bayesian interpretation. Given some prior distribution on nuisance parameters, our test will have the best weighted average powers.

adjustment term that will accommodate the estimation error, rendering our proposed statistic will be asymptotically equivalent to the Likelihood Ratio test even though unknown true $\boldsymbol{\theta}_0$ is replaced by the QMLE under the null.[6] Therefore, it is easily seen that the weak convergence of our test statistic in (3.1) to a Gaussian stochastic process comes directly from Central Limit Theorem for Martingale difference sequences.

We now summarise the asymptotic distributional properties of our proposed test statistics in the following theorem:

Theorem 3.2. *Assume Assumptions 1-3 hold. Under H_0, we have*

$$\Psi_T(\boldsymbol{\lambda}) - \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\xi_t(\boldsymbol{\lambda};\boldsymbol{\theta}_0)}{2} - \mathbf{d}(\boldsymbol{\lambda},\boldsymbol{\theta}_0)' l_t^{(1)}(\boldsymbol{\theta}_0)\right)\right.$$
$$\left. -\frac{1}{2}E_{\theta_0}\left(\frac{\xi_t(\boldsymbol{\lambda};\boldsymbol{\theta}_0)}{2} - \mathbf{d}(\boldsymbol{\lambda},\boldsymbol{\theta}_0)' l_t^{(1)}(\boldsymbol{\theta}_0)\right)^2\right] \to 0 \tag{3.5}$$

uniformly on all compact sets. Furthermore, we have under H_0,

$$\Psi_T(\boldsymbol{\lambda}) \Rightarrow G_1(\boldsymbol{\lambda}),$$

and under H_{1T},

$$\Psi_T(\boldsymbol{\lambda}) \Rightarrow G_2(\boldsymbol{\lambda})$$

where $\Rightarrow$ denotes weak convergence in distribution, $G_1(\boldsymbol{\lambda})$ and $G_2(\boldsymbol{\lambda})$ are both Gaussian processes with different means $-k(\boldsymbol{\lambda},\boldsymbol{\lambda})/2$ and $k(\boldsymbol{\lambda},\boldsymbol{\lambda}_0) - k(\boldsymbol{\lambda},\boldsymbol{\lambda})/2$, and the same covariance $k(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)$ given by

$$k(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) =$$
$$E_{\theta_0}\left\{\left(\frac{\xi_t(\boldsymbol{\lambda}_1;\boldsymbol{\theta}_0)}{2} - \mathbf{d}(\boldsymbol{\lambda}_1,\boldsymbol{\theta}_0)' l_t^{(1)}(\boldsymbol{\theta}_0)\right)\left(\frac{\xi_t(\boldsymbol{\lambda}_2;\boldsymbol{\theta}_0)}{2} - \mathbf{d}(\boldsymbol{\lambda}_2,\boldsymbol{\theta}_0)' l_t^{(1)}(\boldsymbol{\theta}_0)\right)\right\},$$

with $\boldsymbol{\lambda}_0$ being the true value of $\boldsymbol{\lambda}$ under the alternative.

Under Assumption 3, it is easily seen that the Hessian matrix for $\boldsymbol{\theta} = (\mu, \boldsymbol{\theta}_\sigma')'$ with $\boldsymbol{\theta}_\sigma = (\alpha_0, \alpha_1, \gamma_1)'$ is block diagonal. This implies that the GARCH model, (2.1), can be estimated consistently by the two-stage QML estimation

[6] Jensen and Rahbeck (2004) prove under an additional condition of the finite 4th moment of the innovation that the QML estimator is asymptotically normal even in the nonstationary IGARCH case.

since the QMLE of the conditional mean equation has no effect on the conditional variance parameters, $\boldsymbol{\theta}_\sigma$, asymptotically. So we focus on the concentrated conditional log-likelihood function:

$$\ell_t^c = \ell_t^c(\boldsymbol{\theta}_\sigma) \propto -\frac{1}{2}\ln\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right) - \frac{v_t^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)}. \tag{3.6}$$

Then the first and second derivatives are derived as follows:

$$\begin{aligned}
\frac{\partial \ell_t^c}{\partial \alpha_0} &= -\frac{1}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} + \frac{v_t^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2}, \quad (3.7)\\
\frac{\partial \ell_t^c}{\partial \alpha_1} &= -\frac{v_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} + \frac{v_t^2 v_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2},\\
\frac{\partial \ell_t^c}{\partial \gamma_1} &= -\frac{\sigma_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} + \frac{v_t^2 \sigma_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell_t^c}{\partial \alpha_0^2} &= \frac{1}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2} - \frac{v_t^2}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^3}, \quad (3.8)\\
\frac{\partial^2 \ell_t^c}{\partial \alpha_0 \partial \alpha_1} &= \frac{v_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2} - \frac{v_t^2 v_{t-1}^2}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^3},\\
\frac{\partial^2 \ell_t^c}{\partial \alpha_0 \partial \gamma_1} &= \frac{\sigma_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2} - \frac{v_t^2 \sigma_{t-1}^2}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^3},\\
\frac{\partial^2 \ell_t^c}{\partial \alpha_1^2} &= \frac{v_{t-1}^4}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2} - \frac{v_t^2 v_{t-1}^4}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^3},\\
\frac{\partial^2 \ell_t^c}{\partial \alpha_1 \partial \gamma_1} &= \frac{v_{t-1}^2 \sigma_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2} - \frac{v_t^2 v_{t-1}^2 \sigma_{t-1}^2}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^3},\\
\frac{\partial^2 \ell_t^c}{\partial \gamma_1^2} &= \frac{\sigma_{t-1}^4}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^2} - \frac{v_t^2 \sigma_{t-1}^4}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)^3}.
\end{aligned}$$

We describe the detailed testing procedure in what follows: We first obtain the QMLE of $\boldsymbol{\theta} = (\mu, \alpha_0, \alpha_1, \gamma_1)'$, denoted $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\gamma}_1)'$, $\hat{v}_t = y_t - \hat{\mu}$ and $\hat{\sigma}_t^2$ from the null GARCH model, (2.1), and plug them in $\frac{\partial \ell_t^c(\boldsymbol{\theta}_\sigma)}{\partial \boldsymbol{\theta}_\sigma}$ and $\frac{\partial^2 \ell_t^c(\boldsymbol{\theta}_\sigma)}{\partial \boldsymbol{\theta}_\sigma \partial \boldsymbol{\theta}_\sigma'}$ given by (3.7) and (3.8), recursively. The initial value for $\hat{\sigma}_t^2$ is taken to be the estimated unconditional variance as is standard in the literature.

Next, to deal with non-identifiability of $\boldsymbol{\lambda}$ under the null, we need to calculate the summary test statistic, Ψ_{sup} and Ψ_{exp}. For the Ψ_{sup} statistic we can find an analytic solution form when maximizing $\Psi_T(\boldsymbol{\lambda})$ with respect to c^2, which is given by

$$\Psi_{\text{sup}} = \sup_{\{\boldsymbol{\delta}_\sigma,\rho:\|\boldsymbol{\delta}\|=1,\underline{\rho}<\rho<\bar{\rho}\}} \frac{1}{2}\left(\max\left(0, \frac{\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\xi_t^*\left(\boldsymbol{\lambda};\hat{\boldsymbol{\theta}}_\sigma\right)}{\sqrt{T^{-1}\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*}}\right)\right)^2,$$

where

$$\xi_t^*(\boldsymbol{\lambda};\boldsymbol{\theta}_\sigma) =$$
$$\boldsymbol{\delta}_\sigma'\left[\left(\frac{\partial^2\ell_t^c}{\partial\boldsymbol{\theta}_\sigma\partial\boldsymbol{\theta}_\sigma'} + \left(\frac{\partial\ell_t^c}{\partial\boldsymbol{\theta}_\sigma}\right)\left(\frac{\partial\ell_t^c}{\partial\boldsymbol{\theta}_\sigma}\right)'\right) + 2\sum_{s=1}^{t-1}\rho^{(t-s)}\left(\frac{\partial\ell_t^c}{\partial\boldsymbol{\theta}_\sigma}\right)\left(\frac{\partial\ell_t^c}{\partial\boldsymbol{\theta}_\sigma}\right)'\right]\boldsymbol{\delta}_\sigma, \quad (3.9)$$

$\boldsymbol{\delta}_\sigma$ is a three-dimensional vector corresponding to the dimension of $\boldsymbol{\theta}_\sigma$, $\hat{\boldsymbol{\varepsilon}}^*$ is the residual from the OLS regression of $\frac{1}{2}\xi_t^*\left(\boldsymbol{\lambda};\hat{\boldsymbol{\theta}}_\sigma\right)$ on $l_t^{(1)}\left(\hat{\boldsymbol{\theta}}_\sigma\right)$, and $\hat{\boldsymbol{\theta}}_\sigma$ is the QMLE of $\boldsymbol{\theta}_\sigma$ obtained under the null GARCH(1,1) model. Since $\boldsymbol{\Gamma}_T^*$ and $\hat{\boldsymbol{\varepsilon}}^*$ no longer depend on c^2, we need to do grid search over $\boldsymbol{\delta}_\sigma$ and ρ only. To obtain the grid set, $\{\boldsymbol{\delta}_\sigma,\rho : \|\boldsymbol{\delta}_\sigma\| = 1, \underline{\rho} < \rho < \bar{\rho}\}$, we draw $\boldsymbol{\delta}_\sigma$ randomly by

$$\boldsymbol{\delta}_\sigma = \mathbf{z}/\sqrt{\mathbf{z}'\mathbf{z}}, \quad (3.10)$$

where $\mathbf{z}$ is a 3-dimensional Gaussian vector, so that $\boldsymbol{\delta}_\sigma$ is uniform over the unit sphere.

We then draw ρ from an equi-spaced grid over the interval $\left(\underline{\rho},\bar{\rho}\right)$. This procedure will be repeated R times to tabulate the empirical critical values of the Ψ_{sup} statistic. We note in passing that the asymptotic null distribution of our proposed statistic is not pivotal due to its dependence on $\boldsymbol{\theta}$ or $\boldsymbol{\theta}_\sigma$.

The evaluation of the Ψ_{exp} statistic is more involved since we need to pick some prior distributions for all nuisance parameters in $\boldsymbol{\lambda}_\sigma = (c^2, \boldsymbol{\delta}_\sigma', \rho)'$. The most commonly used priors are uniform distributions. But since c^2 is not bounded from above, a uniform prior is not appropriate. We let $C = c^2$ and take an exponential prior for C, which has the pdf of $\tau e^{-\tau C}$. The reason for picking an exponential distribution is that we are able to obtain an analytic form when taking the integral of $\exp(\Psi_T)$ with respect to the distribution for C, denoted $I\Psi_C$, as follows:

$$I\Psi_C = \sqrt{2\pi}\frac{\tau}{\sqrt{\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*/T}}\exp\left[\frac{\left(\boldsymbol{\Gamma}_T^*-\tau\right)^2}{2\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*/T}\right]\Phi\quad\frac{\boldsymbol{\Gamma}_T^*-\tau}{\sqrt{\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*/T}}\Bigg), \quad (3.11)$$

where $\mathbf{\Gamma}_T^* = \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \xi_t^* \left(\boldsymbol{\lambda}; \hat{\boldsymbol{\theta}}_\sigma \right)$ and $\Phi(\cdot)$ is the standard normal cdf. (See Appendix for a derivation of (3.11).) For the other two nuisance parameters $(\boldsymbol{\delta}_\sigma', \rho)'$, we pick uniform distributions over some compact spaces as described above. Then, $\Psi_{\exp}$ is obtained as the average of $I\Psi_C$ over the grid set of $\boldsymbol{\delta}_\sigma$ and ρ. In practice the choice of τ may be arbitrary though we use the same value when generating the critical values and calculating the actual test statistic. We consider the following two choices: (i) Fixed τ: for example, we pick $\tau = 1$ both under the null and the alternative. (ii) Data dependent selection of τ: we suggest to take $\tau = \sqrt{\frac{\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*}{T}}$ under the null. In this case (3.11) simplifies to

$$I\Psi_C = \sqrt{2\pi} \exp \left[\frac{1}{2} \left. \quad \frac{\mathbf{\Gamma}_T^*}{\sqrt{\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*/T}} - 1 \right)^2 \right] \Phi \left. \quad \frac{\mathbf{\Gamma}_T^*}{\sqrt{\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*/T}} - 1 \right).$$

Under the alternative, we plug the average of $\sqrt{\frac{\hat{\boldsymbol{\varepsilon}}^{*\prime}\hat{\boldsymbol{\varepsilon}}^*}{T}}$ evaluated under the null for τ in (3.11). Simulation results (unreported) show that the power performance based on the fixed case with $\tau = 1$ is reasonably close to those based on a data dependent choice. Hence we suggest to use a data dependent choice for the $\Psi_{\exp}$ test.[7]

It has been well-documented that neglecting structural breaks or parameter changes in GARCH models would result in a spurious finding of an extremely high persistence in the unconditional variance. Following Maekawa *et al.* (2004) we show that the (stationary) GARCH process with a structural break will follow the IGARCH process. Suppose that a structural break in α_0 in (2.1) occurs at time t_0,

$$\sigma_t^2 = \left\{ \begin{array}{ll} \alpha_{01} + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 & t < t_0 \\ \alpha_{02} + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 & t \geq t_0 \end{array} \right\}, \ \alpha_{01} \neq \alpha_{02}.$$

Defining $u_t = v_t^2 - \sigma_t^2$, we have

$$v_t^2 = \left\{ \begin{array}{ll} \frac{\alpha_{01}}{1-\gamma_1} + \alpha_1 (1 - \gamma_1 L)^{-1} v_{t-1}^2 + u_t & t < t_0 \\ \frac{\alpha_{02}}{1-\gamma_1} + \alpha_1 (1 - \gamma_1 L)^{-1} v_{t-1}^2 + u_t & t \geq t_0 \end{array} \right\},$$

which can be written as

$$\Delta v_t^2 = \alpha_1 (1 - \gamma_1 L)^{-1} \Delta v_{t-1}^2 + \eta_t + \Delta u_t, \tag{3.12}$$

[7] These additional simulation results will be available upon request.

where

$$\eta_t = \left\{ \begin{array}{cc} \frac{\alpha_{02}-\alpha_{01}}{1-\gamma_1} & t = t_0 \\ 0 & \text{otherwise} \end{array} \right\}.$$

Embedding the simple MS mechanism as

$$\eta_t = \left\{ \begin{array}{cc} \frac{\alpha_{02}-\alpha_{01}}{1-\gamma_1} & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{array} \right\},$$

where p is a small number to represent that this is a relatively rare event, Maekawa *et al.* (2004) shows that (3.12) above shares the same form as an IGARCH(1,1) process. Hillebrand (2005) further shows in a more general context that the estimated sum of the GARCH parameters converges to one when there are parameter changes in the conditional volatility of the process. Again very little research has been done to develop the test procedure which is able to distinguish between these two processes, one being a nonstationary IGARCH process, and the other being a regime-dependent globally stationary process. See also Francq and Zakoïan (2001), Mikosch and Starica (2004) and Smith (2006).

The testing procedure derived above can be easily adapted to develop the optimal test for the null of IGARCH against the alternative of (globally) stationary MS-GARCH. In fact, the procedure remains the same but the calculation is simpler due to the IGARCH restriction, $\alpha_1 + \gamma_1 = 1$, being imposed under the null. Now, there are only two parameters in $\boldsymbol{\theta}_\sigma$, namely, $\boldsymbol{\theta}_\sigma = (\alpha_0, \alpha_1)'$ and thus $\frac{\partial l_t}{\partial \boldsymbol{\theta}_\sigma}$ is two-dimensional. It is easily seen that the proposed IGARCH null test is consistent against the alternative of (globally) stationary MS-GARCH processes, which include one-time structural break as a special case. Therefore, the proposed test may be used as a generic test for detecting structural breaks or parameter changes in GARCH models.

4. Monte Carlo Simulation

In this section we undertake a Monte Carlo investigation of the finite sample performance of our proposed tests under Assumption 3 for the null of GARCH or IGARCH process against stationary MS-GARCH process, denoted $\Psi_{\sup}^{GARCH}$, $\Psi_{\sup}^{IGARCH}$, $\Psi_{\exp}^{GARCH}$, and $\Psi_{\exp}^{IGARCH}$ respectively.

To evaluate the performance of the $\Psi_{\sup}^{GARCH}$ and $\Psi_{\exp}^{GARCH}$ tests we first consider the following data generating process under the null hypothesis:

$$\left\{ \begin{array}{c} y_t = \mu + v_t \\ v_t = \varepsilon_t \sigma_t,\ \varepsilon_t \sim iid\mathcal{N}(0,1) \\ \sigma_t^2 = \alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \end{array} \right. \tag{4.1}$$

where we set $\mu = 1$, $\alpha_0 = 0.1$, $\alpha_1 = 0.1$, $\gamma_1 = 0.8$ and the unconditional variance is equal to 1. We consider two regimes under the alternative and let the unobserved Markov Chain, S_t, take binary values with transition probabilities given by $p = \Pr(S_t = 0|S_{t-1} = 0)$ and $q = \Pr(S_t = 1|S_{t-1} = 1)$, where $S_t = 0$ corresponds to the regime where the parameters take the same values as under the null hypothesis whilst $S_t = 1$ corresponds to the other regime where the parameters switch. The stationary distribution for S_t is then given by

$$\Pr(S_t = 0) = \frac{1-q}{2-p-q};\ \Pr(S_t = 1) = \frac{1-p}{2-p-q}.$$

To compute the $\Psi_{\text{sup}}^{GARCH}$ and $\Psi_{\text{exp}}^{GARCH}$ test statistics over the grid set, $\{\boldsymbol{\delta},\rho : \|\boldsymbol{\delta}\| = 1, \underline{\rho} < \rho < \bar{\rho}\}$ as described in Section 3, we use 40 draws for $\boldsymbol{\delta}$ from the unit sphere and 30 draws for ρ from an equi-spaced grid over the interval $(\underline{\rho}, \bar{\rho}) = (0, 0.98)$.[8] We set the sample size to 400 and use 1,000 replications to calculate the empirical critical values and then the size-corrected powers.[9]

We consider two experiments denoted Experiment 1A and Experiment 1B, where α_{00}, α_{10} and γ_{10} are GARCH parameters in the null regime while α_{01}, α_{11} and γ_{11} GARCH parameters in an alternative regime. Experiment 1A allows the switching only in α_0 and sets $\alpha_{00} = 0.1$ and $\alpha_{01} = (0.1, 0.2, 0.5, 1, 2, 5, 10, 20)$, where the first element ($\alpha_{01} = 0.1$) corresponds to the null. In Experiment 1B, we allow that the switching happens in α_0, α_1 and γ_1, simultaneously. Under the null hypothesis, we set $\alpha_{10} + \gamma_{10} = 0.9$. To investigate the power of the tests we switch this sum to $\{0.9, 0.7, 0.5, 0.3, 0.1, 0.05\}$, respectively. For convenience we change the values of α_{01}, α_{11} and γ_{11}, proportionally. For example, when the sum is 0.7, we make $\alpha_{01} = \frac{0.1}{0.9} \times 0.7$, $\alpha_{11} = \frac{0.1}{0.9} \times 0.7$ and $\gamma_{11} = \frac{0.8}{0.9} \times 0.7$ such that $\alpha_{11} + \gamma_{11} = 0.7$ and similarly for other values.

The results for Experiment 1A and Experiment 1B are summarised in Table 1. The power performances of both $\Psi_{\text{sup}}^{GARCH}$ and $\Psi_{\text{exp}}^{GARCH}$ tests are reasonably satisfactory in both experiments. As the alternative gets more distant from the null, the power increases monotonically. Looking at the upper panel for Experiment 1A we also find that the power depends on the respective transition probabilities, namely the power increases as regimes are more absorbing and persistent. Considering that the regimes have been found to be very persistent by most empirical studies, *e.g.* Chaudhuri and Klassen (2001), we may

[8] The range of ρ is determined by noting that $\rho = p + q - 1$ and $(p, q) \in (0.5, 0.99)$. We find that an increase in the number of draws does not have a significant impact on the outcome.

[9] There are substantial small sample biases in estimating GARCH parameters, *e.g.* Lumsdaine (1995) and Hillebrand (2005). As the sample size increases, these estimates become more accurate.

regard the case with $(p, q) = (0.98, 0.98)$ as a benchmark. On the other hand this pattern is less obvious in Experiment 1B when all three parameters are switching simultaneously. Moreover, exponential test is performing slightly better than sup test in Experiment 1A when the alternative is distant from the alternative whereas a similar pattern is not observed in Experiment 1B.

Table 1. Power of $\Psi^{GARCH}_{\text{sup}}$ and $\Psi^{GARCH}_{\text{exp}}$ tests against MS-GARCH Models

Experiment 1A												
(p, q)	$(0.98, 0.98)$				$(0.9, 0.9)$				$(0.98, 0.9)$			
size	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
α_0	$\Psi^{GARCH}_{\text{sup}}$		$\Psi^{GARCH}_{\text{exp}}$		$\Psi^{GARCH}_{\text{sup}}$		$\Psi^{GARCH}_{\text{exp}}$		$\Psi^{GARCH}_{\text{sup}}$		$\Psi^{GARCH}_{\text{exp}}$	
0.1	6.2	10.0	4.8	11.9	5.2	10.6	5	12.1	6.2	10.0	4.6	11.5
0.2	9.1	15.9	7.4	16.7	7.8	12.0	6	18	6.7	12.2	5.9	15.7
0.5	20.3	30.5	19.9	39.6	15.0	23.6	13.8	30.2	15.8	25.8	18.7	36.2
1	32.0	45.4	35.1	57.6	26.9	40	23.4	45.8	34.7	47.2	42.1	63
2	53.4	66.9	61.2	78.8	39.2	52.1	36.9	62.2	63.4	76.0	68	84
5	72.9	83.1	81.9	90.9	47.3	60.5	49.3	71.3	88.9	94.9	92.1	98.3
10	81.6	88.5	89	95.9	47.4	60.7	51.9	71.9	95.9	98.4	98.3	99.1
20	86.6	91.7	91.7	96.4	47.6	61.0	59.4	76.7	99.3	99.9	99.7	100

Experiment 1B												
(p, q)	$(0.98, 0.98)$				$(0.9, 0.9)$				$(0.98, 0.9)$			
size	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$\alpha_1+\gamma_1$	$\Psi^{GARCH}_{\text{sup}}$		$\Psi^{GARCH}_{\text{exp}}$		$\Psi^{GARCH}_{\text{sup}}$		$\Psi^{GARCH}_{\text{exp}}$		$\Psi^{GARCH}_{\text{sup}}$		$\Psi^{GARCH}_{\text{exp}}$	
0.9	6.5	10.1	4.8	11.9	6.4	10.5	5	12.1	6.5	10.1	4.6	11.5
0.7	23.5	33.7	15.6	27.1	18.9	27.2	19.6	31.9	13.9	19.7	13.6	26.2
0.5	57.6	66.8	27.8	35.6	72	80.1	62.9	73.7	41.9	50.6	44.9	61.9
0.3	85.3	90.6	26.5	33	98.1	99	89.2	92.8	74.4	80.9	75.5	82.7
0.1	97	98.8	37.4	42.1	100	100	95.1	96.4	90.8	93	82.3	86.8
0.05	97.4	98.8	40.4	43	100	100	95.4	96.5	91.4	93.3	85.5	88.9

We consider Experiment 2 where the sum of GARCH parameters is less than unity in one regime but greater than or equal to 1 in the other regime; namely, $\alpha_{10} + \beta_{10} < 1$ and $\alpha_{11} + \beta_{11} \geq 1$. In this case, one regime is stationary and the other regime is nonstationary. So this case covers an interesting case where the process is locally nonstationary but globally stationary. The condition for global stationarity is given by $\kappa(\mathbf{M}) < 1$ (see Francq and Zakoïan (2001)), where $\kappa(\cdot)$ denotes the largest eigenvalue of a matrix $\mathbf{M}$ defined by

$$\mathbf{M} = \begin{bmatrix} p(\alpha_{10}+\gamma_{10}) & 0 & (1-q)(\alpha_{10}+\gamma_{10}) & 0 \\ p\alpha_{11} & p\gamma_{11} & (1-q)\alpha_{11} & (1-q)\gamma_{11} \\ (1-p)\gamma_{10} & (1-p)\alpha_{10} & q\gamma_{10} & q\alpha_{10} \\ 0 & (1-p)(\alpha_{11}+\gamma_{11}) & 0 & q(\alpha_{11}+\gamma_{11}) \end{bmatrix}.$$

Here we select $\mu = 1$, $\alpha_0 = 0.2$, $\alpha_1 = 0.05$, $\gamma_1 = 1 - \alpha_1 = 0.95$ under the null. We use the same design as in Experiment 1B under the alternative; we make the sum, $\alpha_1 + \gamma_1$ switch to $\{1, 0.9, 0.7, 0.5, 0.3, 0.1\}$, and change the values of α_0, α_1 and γ_1 proportionally as before. The critical values under the null of IGARCH are first evaluated and then the size-corrected powers are assessed. The results reported in Table 2 demonstrate that the power performance of the $\Psi_{\text{sup}}^{GARCH}$ test is still reasonably satisfactory. As the alternative gets more distant from the null, the power increases monotonically.[10] But, it does not show that the power depends on the respective transition probabilities in any meaningful sense unlike the previous case.

Table 2. Power of $\Psi_{\text{sup}}^{IGARCH}$ test against MS-GARCH Models

Experiment 2						
(p,q)	$(0.98, 0.98)$		$(0.9, 0.9)$		$(0.98, 0.9)$	
size	5%	10%	5%	10%	5%	10%
$\alpha_1+\gamma_1$	$\Psi_{\text{sup}}^{IGARCH}$		$\Psi_{\text{sup}}^{IGARCH}$		$\Psi_{\text{sup}}^{IGARCH}$	
1	5.3	9.8	5.3	9.6	5.3	9.8
0.9	11.8	23.5	10.2	19.6	10.3	18.4
0.7	43.0	62.3	37.8	59.5	44.5	63.1
0.5	66.3	81.1	79.8	92.2	76.4	88.3
0.3	80.8	91.3	98.4	99.8	92.9	96.4
0.1	86.5	94.5	100	100	94.5	98.2

We now turn to a more practical situation, denoted Experiment 3, where the GARCH parameters, of α_0, α_1 and γ_1 under the null hypothesis are unknown and thus need to be estimated by QML in order to compute the empirical critical values. We generate the data by setting $\alpha_1 = 0.1$ and $\gamma_1 = 0.7$ but

[10] We do not consider the $\Psi_{\exp}$ test since its value will become extremely large whenever the GARCH estimates are close to the boundary, which makes the evaluation of critical values unreliable.

switching α_0 between 0.2 and 2 with transition probabilities 0.98 and 0.98.[11] The rejection probabilities of our proposed tests and the conventional (robust) Wald test for the null of $\alpha_1 + \gamma_1 = 1$ are presented in Table 3. As expected the robust Wald test rejects the null of IGARCH only 1.9% of the time. On the other hand our proposed tests correctly and strongly reject the null of GARCH or IGARCH in favor of Markov Switching alternatives. This clearly demonstrates that neglecting the parameter changes in GARCH models would result in a spurious finding of IGARCH estimates.

Table 3. Critical Values and p-Values for Experiment 3

	5%	10%	p-value
$\Psi_{\text{sup}}^{GARCH}$	2.986	2.336	0.009
$\Psi_{\text{sup}}^{IGARCH}$	2.493	1.858	0.025
$\Psi_{\text{exp}}^{GARCH}$	1.780	1.256	0.004

In practice it is interesting to see how robust our proposed test based on the QML procedure is to the misspecified error distribution. In particular, to investigate the finite sample performance of our proposed tests in presence of fat-tailed innovations, we consider an additional simulation design in which we generate the data using similar parameter specifications used above but with a main difference that the innovation follows a t-distribution with r degrees of freedom. The kurtosis of Student t random variable equals $3(r-2)/(r-4)$, implying that the kurtosis is uniquely determined by r. We set $r = 7$, and then the kurtosis is equal to 5.

We now pick the DGP similar to Experiment 1B. Namely, α_0, α_1 and γ_1 change simultaneously and proportionally when the sum of α_1 and γ_1 take the values of $\{0.9, 0.7, 0.5, 0.3, 0.1, 0.05\}$. First, we investigate the robustness of our proposed tests to misspecification. To this end we apply the $\Psi_{\text{sup}}^{GARCH}$ test obtained assuming the normally distributed errors as described above. Panel A of Table 4 summarises the small sample performance of the $\Psi_{\text{sup}}^{GARCH}$ test. The results demonstrate that the $\Psi_{\text{sup}}^{GARCH}$ test is still able to detect the switching in parameters reasonably well even in presence of misspecification.

[11] Qualitatively similar results have been obtained for different values of transition probabilities.

Table 4. Power of $\Psi_{\sup}^{GARCH}$ test against MS-GARCH Models with t_7 Errors

Panel A: Misspecified Model						
(p,q)	$(0.98, 0.98)$		$(0.9, 0.9)$		$(0.98, 0.9)$	
size	5%	10%	5%	10%	5%	10%
$\alpha_1+\gamma_1$	$\Psi_{\sup}^{GARCH}$		$\Psi_{\sup}^{GARCH}$		$\Psi_{\sup}^{GARCH}$	
0.9	5.5	10.4	4	8.3	5.3	9
0.7	11.4	20	9.9	18.1	7.3	13.7
0.5	24.9	37.3	28.8	44.2	17	27.2
0.3	41.7	56.2	64.1	78.1	31.5	44.9
0.1	61	73.1	94.4	98.1	51.1	65.2
0.05	66.6	79.6	96.8	99.1	61.5	72.9

Panel B: Correctly Specified Model						
(p,q)	$(0.98, 0.98)$		$(0.9, 0.9)$		$(0.98, 0.9)$	
size	5%	10%	5%	10%	5%	10%
$\alpha_1+\gamma_1$	$\Psi_{\sup}^{tGARCH}$		$\Psi_{\sup}^{tGARCH}$		$\Psi_{\sup}^{tGARCH}$	
0.9	5	11.3	4.2	11.1	5.1	11.4
0.7	10	16.7	10.1	15.9	7.9	16.8
0.5	43.9	51.9	32.1	38.3	33.5	48.2
0.3	67.2	77.9	39.3	50.5	53.1	66.6
0.1	79.5	87.2	59	69.6	69.3	73.2
0.05	80.4	88.8	68.1	77.2	72.6	80.7

But the power loss due to misspecification is non-negligible especially when the degree of switching is relatively small. It is therefore worth re-examining the test performance based on the correctly specified model with t-distribution.[12] Assuming a t-distribution, the conditional log-likelihood function for v_t in (4.1) is now given by

$$
\begin{aligned}
&l_t(\boldsymbol{\theta}_\sigma) \propto \\
&\ln\Gamma\left(\frac{r+1}{2}\right) - \frac{1}{2}\ln(r-2) - \ln\Gamma\left(\frac{r}{2}\right) - \frac{1}{2}\ln\left(\alpha_0+\alpha_1 v_{t-1}^2+\gamma_1\sigma_{t-1}^2\right) \\
&\quad -\frac{r+1}{2}\ln\left(1+\frac{v_t^2}{(r-2)(\alpha_0+\alpha_1 v_{t-1}^2+\gamma_1\sigma_{t-1}^2)}\right) \qquad (4.2)
\end{aligned}
$$

[12]In practice, the GARCH estimation has been performed using either normal or t-distribution and the final preferred specification is selected by various pre- and post-diagnostics.

where $\Gamma(\cdot)$ denotes a Gamma function. Since we now have one more parameter, r, we set $\boldsymbol{\theta}_\sigma = (\alpha_0, \alpha_1, \gamma_1, r)'$. The first derivatives are obtained by

$$
\begin{aligned}
\frac{\partial l_t}{\partial \alpha_0} &= -\frac{1}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} + \frac{r+1}{2}\frac{\vartheta_t}{1+\vartheta_t}\frac{1}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)} \\
\frac{\partial l_t}{\partial \alpha_1} &= -\frac{v_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} + \frac{r+1}{2}\frac{\vartheta_t}{1+\vartheta_t}\frac{v_{t-1}^2}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} \\
\frac{\partial l_t}{\partial \gamma_1} &= -\frac{\sigma_{t-1}^2}{2\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} + \frac{r+1}{2}\frac{\vartheta_t}{1+\vartheta_t}\frac{\sigma_{t-1}^2}{\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)} \\
\frac{\partial l_t}{\partial r} &= \frac{1}{2}\left[\quad \left(\frac{r+1}{2}\right) - \frac{1}{(r-2)} - \quad \left(\frac{r}{2}\right) - \ln(1+\vartheta_t) + \frac{(r+1)\vartheta_t}{(r-2)(1+\vartheta_t)}\right]
\end{aligned}
$$

where $\psi(\cdot)$ is the digamma function; $\vartheta_t = v_t^2 / \left\{(r-2)\left(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2\right)\right\}$ is defined for notational simplicity. Similarly, the second derivatives are as follows:

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial \alpha_0^2} &= \frac{1}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)^2}\left[\frac{1}{2} - \frac{r+1}{2}\frac{(2+\vartheta_t)\vartheta_t}{(1+\vartheta_t)^2}\right] \\
\frac{\partial^2 l_t}{\partial \alpha_0 \partial \alpha_1} &= \frac{v_{t-1}^2}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)^2}\left[\frac{1}{2} - \frac{r+1}{2}\frac{(2+\vartheta_t)\vartheta_t}{(1+\vartheta_t)^2}\right] \\
\frac{\partial^2 l_t}{\partial \alpha_0 \partial \gamma_1} &= \frac{\sigma_{t-1}^2}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)^2}\left[\frac{1}{2} - \frac{r+1}{2}\frac{(2+\vartheta_t)\vartheta_t}{(1+\vartheta_t)^2}\right] \\
\frac{\partial^2 l_t}{\partial \alpha_0 \partial r} &= \frac{\vartheta_t}{2(1+\vartheta_t)}\frac{1}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)} - \frac{r+1}{2}\frac{\vartheta_t^2}{(1+\vartheta_t)^2 v_t^2} \\
\frac{\partial^2 l_t}{\partial \alpha_1^2} &= \frac{v_{t-1}^4}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)^2}\left[\frac{1}{2} - \frac{r+1}{2}\frac{(2+\vartheta_t)\vartheta_t}{(1+\vartheta_t)^2}\right] \\
\frac{\partial^2 l_t}{\partial \alpha_1 \partial \gamma_1} &= \frac{v_{t-1}^2 \sigma_{t-1}^2}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)^2}\left[\frac{1}{2} - \frac{r+1}{2}\frac{(2+\vartheta_t)\vartheta_t}{(1+\vartheta_t)^2}\right] \\
\frac{\partial^2 l_t}{\partial \alpha_1 \partial r} &= \frac{\vartheta_t}{2(1+\vartheta_t)}\frac{v_{t-1}^2}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)} - \frac{r+1}{2}\frac{\vartheta_t^2}{(1+\vartheta_t)^2}\frac{v_{t-1}^2}{v_t^2} \\
\frac{\partial^2 l_t}{\partial \gamma_1^2} &= \frac{\sigma_{t-1}^4}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)^2}\left[\frac{1}{2} - \frac{r+1}{2}\frac{(2+\vartheta_t)\vartheta_t}{(1+\vartheta_t)^2}\right] \\
\frac{\partial^2 l_t}{\partial \gamma_1 \partial r} &= \frac{\vartheta_t}{2(1+\vartheta_t)}\frac{\sigma_{t-1}^2}{(\alpha_0 + \alpha_1 v_{t-1}^2 + \gamma_1 \sigma_{t-1}^2)} - \frac{r+1}{2}\frac{\vartheta_t^2}{(1+\vartheta_t)^2}\frac{\sigma_{t-1}^2}{v_t^2} \\
\frac{\partial^2 l_t}{\partial r^2} &= \frac{1}{4}\varphi\left(\frac{r+1}{2}\right) + \frac{1}{2(r-2)^2} - \frac{1}{4}\varphi\left(\frac{r}{2}\right) + \frac{\vartheta_t(\vartheta_t r - 5\vartheta_t - 6)}{2(r-2)^2(1+\vartheta_t)^2}
\end{aligned}
$$

where $\varphi(\cdot)$ is trigamma function.

The testing procedure is similar to those described in Section (3). We first obtain MLE of $\boldsymbol{\theta} = (\mu, \alpha_0, \alpha_1, \gamma_1, r)'$ from t-$GARCH$ distribution. Then, we plug them into $\dfrac{\partial l_t(\boldsymbol{\theta}_\sigma)}{\partial \boldsymbol{\theta}_\sigma}$ and $\dfrac{\partial^2 l_t(\boldsymbol{\theta}_\sigma)}{\partial \boldsymbol{\theta}_\sigma \partial \boldsymbol{\theta}_\sigma'}$ and construct $\xi_t^*(\boldsymbol{\lambda}; \boldsymbol{\theta}_\sigma)$ using (3.9) as before. There are two cases: (i) Parameter switching only occurs in GARCH parameters, namely, α_0, α_1 and γ_1 only but the degree of freedom parameter r is not subject to switching and (ii) All t-$GARCH$ parameters switch. For case (i) we generate $\boldsymbol{\delta}$ to be a 4-dimensional vector, with the elements corresponding to α_0, α_1 and γ_1 to be uniform over the unit sphere and the last element of $\boldsymbol{\delta}$ to be set to 0. For a more general case (ii) we also allow for switching in r, implying that the kurtosis also switches in different regimes. In this case, we generate $\boldsymbol{\delta}$ to be a 4-dimensional vector which is uniform over the unit sphere. We then follow the same testing procedure as described in Section (3).

The simulation results based on the correctly specified log-likelihood function, (4.2) for the DGP with the t-distributed innovations are presented in Panel B of Table 4. To simplify a simulation design and to match with the same parameter specification as used in Experiment B, we allow the switching to takes place only in a subset of parameters, $(\alpha_0, \alpha_1, \gamma_1)$, which corresponds to Case (i). But we still project $\xi_t^*(\boldsymbol{\lambda}; \boldsymbol{\theta}_\sigma)$ on the four dimensional vector, $\dfrac{\partial l_t(\boldsymbol{\theta}_\sigma)}{\partial \boldsymbol{\theta}_\sigma}$ and derive the residual vector, $\hat{\boldsymbol{\varepsilon}}^*$. Comparing with the results in Panel A, there is a significant power boost from using the tests derived from the correctly-specified model.

5. Empirical Illustration

The apparent IGARCH behavior in the conditional volatility of the stock returns has become an awkward puzzle for economists. The difficulty of rejecting an IGARCH implies an implausible conclusion that any structural shocks would have a permanent impact on persistence of the conditional volatility in the future. However, it is quite plausible that structural breaks or regime shifts in financial assets markets are likely to lead to these spurious findings. We investigate this issue by applying our proposed tests.

The East Asian financial crises was a financial crisis starting from July 1997 in Thailand and affected currencies, stock markets, and other asset prices during 1997-98. In particular, Indonesia, South Korea and Thailand were most affected by the crisis while Malaysia and the Philippines were also hit. Since then there has been a surge of interest in analysing the banking and currency crises in emerging markets using the regime-switching models in both conditional mean and conditional volatility, *e.g.* Brunetti et al. (2003). Chaudhuri and Klassen (2001) suggest to estimate the two-regime switching GARCH models for the above five countries and find: (i) The GARCH estimates even in the high volatility regime are less pronounced than implied by the standard GARCH model, where the sum of the estimated GARCH parameters are close to one; (ii) Both low and high volatility regimes are very persistent, namely, the average estimates of probability staying in the same regime are respectively 0.992 and 0.989. So both regimes persist for about two years; (iii) All stock indices were in the high volatility regime in the second half of 1997 and in 1998, the crisis period; (iv) The stock returns were still in the high-volatility regime in June 2000 (with the exception of the Philippines), and the increment in volatility caused by the crisis has been removed only by about 60%.

In almost all studies applying the regime-switching GARCH models, however, there has been no specification tests for the null of GARCH models. We therefore apply our proposed tests to the weekly percentage returns computed from the stock price indices for South Korea, Thailand, Malaysia and the Philippines from January 5, 1989 to June 1, 2000 (a total of 596 observations) and for Indonesia from October 11, 1990 to June 1, 2000 (a total of 504 observations). The data is collected from the International Finance Corporation's Emerging Market Database.[13] The estimation and test results based on the standard normal errors are summarised in Table 5 below. We find that GARCH parameters, $\hat{\alpha}_1$ and $\hat{\gamma}_1$, are all statistically significant, but more importantly, the sum is close to unity, all exceeding 0.97 for five countries. The robust Wald test is not able to reject the null hypothesis of IGARCH. Viewed through the "eyes" of standard GARCH analysis, this would imply that there is a permanent volatility persistence in the individual stock returns.

Next, we use the $\Psi_{\sup}^{IGARCH}$ test that tests the null of IGARCH against an alternative of (globally) stationary GARCH model and find that there is little support for the null for all five cases except for Korea. Therefore, we may conclude that the IGARCH finding is spurious due to the mis-specification caused by the neglect of structural breaks such as the Asian Crises in GARCH models, a finding consistent with the previous studies, *e.g.* Caporale, Pittis and Spagnolo (2003) and Hillebrand (2005). Finally, to examine whether the MS-GARCH models adopted by Chaudhuri and Klassen (2001) are accepted by the data, we also apply the $\Psi_{\sup}^{GARCH}$ test and find strong evidence in favor of the of MS-GARCH model for all five countries. Overall these findings clearly support our conclusion that neglecting the parameter changes in GARCH models would

[13] See Chaudhuri and Klassen (2001) for details. We are grateful to Klassen who provides the data.

result in the spurious finding of IGARCH estimates.

Table 5. Estimation and Test Results for Five East Asian Stock Returns

	Indonesia	South Korea	Malaysia	Philippines	Thailand
μ	0.29 (.150)	0.003 (.145)	0.27 (.104)	0.092 (.159)	0.28 (.147)
α_0	0.28 (.193)	0.42 (.213)	0.18 (.137)	0.446 (.304)	0.56 (.303)
α_1	0.10 (.044)	0.086 (.026)	0.11 (.038)	0.048 (.020)	0.11 (.052)
γ_1	0.89 (.042)	0.89 (.028)	0.89 (.035)	0.926 (.035)	0.87 (.048)
$Wald$	0.28 [.597]	1.77 [.184]	0.10 [.747]	1.62 [.204]	0.98 [.322]
$\Psi_{\text{sup}}^{GARCH}$	5.58 [.057]	4.41 [.026]	10.72 [.001]	8.58 [.002]	4.95 [.016]
$\Psi_{\text{sup}}^{IGARCH}$	4.46 [.010]	1.74 [.152]	5.01 [.005]	7.48 [.003]	4.49 [.013]
$\Psi_{\text{exp}}^{GARCH}$	5.29 [0.010]	3.53 [.015]	112.23 [.002]	28.39 [.001]	5.69 [.007]
$\Psi_{\text{exp}}^{IGARCH}$	2.24 [0.056]	1.36 [.162]	8.28 [.006]	33.58 [.003]	2.56 [.036]

Notes: The figures inside $(\cdot)$ indicate standard errors. $Wald$ is the robust Wald statistic for the null of IGARCH against the alternative of stationary GARCH, $\Psi_{\text{sup}}^{GARCH}$ and $\Psi_{\text{exp}}^{GARCH}$ are the statistics for the null of GARCH against the alternative of MS-GARCH, and $\Psi_{\text{sup}}^{IGARCH}$ and $\Psi_{\text{exp}}^{IGARCH}$ are the statistics for the null of IGARCH against the alternative of (globally stationary) MS-GARCH. The values insider $[\cdot]$ stand for the p-value of the statistic. The p-values for all our four tests are obtained using the bootstrap distribution of the test statistics with 1000 replications.

6. Concluding Remarks

The investigation of structural breaks or parameter changes in conjunction with nonlinear regime-switching modelling has recently assumed a prominent role in GARCH modelling. It is clear that a stable nonlinear process can be misclassified as a nonstationary IGARCH process when structural breaks or parameter changes are neglected. This can be further misleading both in value-at-risk and forecasting analysis. In this paper we have proposed a simple optimal test procedure that is designed to have a maximal power to detect MS-GARCH mechanisms. Our proposed tests are shown to have a good power in small samples, and an empirical application clearly shows the potential of our approach.

As is always the case when working with nonlinear models, there are several generalisations that could be analysed in future work. *First*, our testing procedure can be easily extended to allow for switching in both the conditional mean and the conditional variance equations jointly. An interesting application may be the error correction model with GARCH errors, where both error correction and GARCH parameters are subject to the regime-dependent parameter changes. *Secondly*, it could be worth conducting an integral Monte Carlo study in order to investigate the relative finite sample performance of our proposed test *vis-a-vis* existing misspecification tests against a broad class of alternative asymmetric and/or structural break GARCH models. *Finally*, given the strength of evidence in favor of the alternative of MS-GARCH models, it would be interesting to estimate the MS-GARCH models under the alternative. Currently, there are two main difficult obstacles: first, estimation of MS-GARCH models is complicated computationally due to the path-dependence, though there are some algorithms available, *e.g.* Gray (1996), Dueker (1997) and Hass *et al.* (2004). Second, the rejection of the null provides no information about the number of regimes under the alternative. We conjecture that these issues can be successfully handled by using the MCMC approach as suggested for example by Tsay (2002).

A. Appendix

A.1. Derivation of (3.11)

$$
\begin{aligned}
I\Psi_C &\equiv \int_0^\infty \exp\left(C\cdot\mathbf{\Gamma}_T^* - \frac{C^2}{2T}\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*\right)\times\tau\exp\left(-\tau C\right)dC \\
&= \int_0^\infty \tau\exp\left[C\left(\mathbf{\Gamma}_T^*-\tau\right) - C^2\frac{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}{2T}\right]dC \\
&= \tau\exp\left[\frac{\left(\mathbf{\Gamma}_T^*-\tau\right)^2}{2\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}\right]\int_0^\infty \exp\left[-\frac{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}{2T}\left(C-\frac{\mathbf{\Gamma}_T^*-\tau}{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}\right)^2\right]dC \\
&= \tau\exp\left[\frac{\left(\mathbf{\Gamma}_T^*-\tau\right)^2}{2\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}\right]\sqrt{\frac{2\pi T}{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}}\int_0^\infty \frac{\exp\left[-\left(C-\frac{\mathbf{\Gamma}_T^*-\tau}{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}\right)^2/\left(2T/\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*\right)\right]}{\sqrt{\left(2\pi T/\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*\right)}}dC
\end{aligned}
$$

$$= \tau \exp\left[\frac{(\mathbf{\Gamma}_T^* - \tau)^2}{2\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}\right]\sqrt{\frac{2\pi T}{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}}\left[1 - \Phi\left(-\frac{\frac{\mathbf{\Gamma}_T^* - \tau}{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}/T}}{\sqrt{T/\hat{\varepsilon}^{*\prime}\hat{\varepsilon}}}\right)\right]$$

$$= \sqrt{2\pi}\frac{\tau}{\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}} \exp\left[\frac{(\mathbf{\Gamma}_T^* - \tau)^2}{2\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}\right]\Phi\left(\frac{\mathbf{\Gamma}_T^* - \tau}{\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*/T}}\right). \blacksquare$$

References

[1] Andrews, D.W.K. and W. Ploberger (1994), "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica* 62, 1383-1414.

[2] Andrews, D.W.K. and W. Ploberger (1995), "Admissibility of the Likelihood Ratio Test when a Nuisance Parameter Is Present Only Under the Alternative", *The Annals of Statistics* 23, 1609-1629.

[3] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-327.

[4] Brunetti, C., R.S. Mariano, C. Scotti and A.H.H. Tan (2003), "Markov Switching GARCH Models of Currency Crises in Southeast Asia", *mimeo.*, University of Pennsylvania.

[5] Busch, T. (2004), "A Robust LR test for the GARCH models", *mimeo.*, University of Aarhus.

[6] Cai, J. (1994), "A Markov Model of Unconditional Variance in ARCH", *Journal of Business and Economic Statistics* 12, 309-316.

[7] Caporale, G. M, N. Pittis and N. Spagnolo (2003), "IGARCH models and structural breaks", *Applied Economics Letters*, Vol 10, Issue 12, 765-768.

[8] Carrasco, M., L. Hu and W. Ploberger (2005), "Optimal Test for Markov Switching", *mimeo.*, University of Rochester.

[9] Chan, F. and M. McAleer (2002), "Maximum Likelihood Estimation of STAR and STAR-GARCH Models: Theory and Monte Carlo Evidence", *Journal of Applied Econometrics* 17, 509-534.

[10] Chaudhuri, K. and F. Klaassen (2001), "Have East Asian Stock Markets Calmed Down? Evidence from a Regime-Switching Model", *mimeo.*, University of Amsterdam.

[11] Chesher, A. (1984), “Testing for Neglected Heterogeneity”, *Econometrica* 52, 865-872.

[12] Cox, D.R. and D. Miller (1965), *The Theory of Stochastic Processes.* Methuen & Co. Ltd.: London.

[13] Davies, R.B. (1987), “Hypothesis Testing When a Nuisance Parameter Is Present Under the Alternative”, *Biometrika* 74, 33-43.

[14] Dueker, M. (1997), “Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility”, *Journal of Business and Economic Statistics* 15, 26-34.

[15] Engle, R.F. and V.K. Ng (1993), “Measuring and Testing the Impact of News on Volatility”, *Journal of Finance* 48, 1749-1777.

[16] Francq, C. and J.M. Zakoian (2001), “Stationary of Multivariate Markov-Switching ARMA Models”, *Journal of Econometrics* 102, 339-364.

[17] Garcia, R. (1998), “Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models”, *International Economic Review* 39, 763-788.

[18] Glosten, L., R. Jagannathan and D. Runkle (1993), “On the Relation Between Expected Value and the Volatility of Nominal Excess Returns on Stocks”, *Journal of Finance* 48, 1779-1901.

[19] Gray, S. (1996), “Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process”, *Journal of Financial Economics* 42, 27-62.

[20] Haas, M., S. Mittnik and M. Paolella (2004), “A New Approach to Markov Switching GARCH Models”, *Journal of Financial Econometrics* 2, 493-530.

[21] Hagerud, G.E. (1997), “A Smooth Transition ARCH Model for Asset Returns”, *SSE/EFI Working Paper Series in Economics and Finance*, No.162.

[22] Hamilton, J.D. (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”, *Econometrica* 57, 357-384.

[23] Hamilton, J. D and R. Susmel (1994), "ARCH and Changes in Regime", *Journal of Econometrics* 64, 307-333.

[24] Hansen, B. E. (1996), "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis", *Econometrica 64*, 414–30.

[25] Halunga, A. and C.D. Orme (2005), "Misspecification Tests for GARCH Models", *mimeo.*, University of Manchester.

[26] Hillebrand, E. (2005), "Neglecting Parameter Changes in GARCH Models", *Journal of Econometrics*, 129, 121-138.

[27] Jensen, S.T. and A. Rahbek (2004), "Asymptotic Normality of the QMLE Estimator of ARCH in the Nonstationary Case", *Econometrica* 72, 641-646.

[28] Jensen, S.T. and A. Rahbek (2004), "Asymptotic Inference for Nonstationary GARCH", Econometric Theory, Vol. 20: 1203-1226

[29] Lamoureux, C.G. and W.D. Lastrapes (1990), "Persistence in Variance, Structural Change and the GARCH Model", *Journal of Business and Economic Statistics* 8, 225–234.

[30] Lee, S., Y. Tokutsu and K. Maekawa (2003), "The CUSUM Test for Parameter Change in GARCH(1,1) Models", *mimeo.*, Seoul National University.

[31] Lundbergh, S. and T. Teräsvirta (2002), "Evaluating GARCH models", *Journal of Econometrics* 110, 417-435.

[32] Maekawa, K., S. Lee and Y. Tokutsu (2004), "A Note on Volatility Persistence and Structural Changes in GARCH Models", *mimeo.*, Hiroshima University.

[33] Mikosch, T. and C. Starica (2004), "Nonstationarities in Financial Time Series, the Long-range Dependence and the IGARCH Effects", *Review of Economics and Statistics*, 86, 378-390.

[34] Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Econometrica* 57, 1361-1401.

[35] Santana E. (1995), "Quadratic ARCH Models", *Review of Economic Studies*, Vol.62: 636-661.

[36] Smith, D. (2006), "Testing for Structural Breaks in GARCH Models", *mimeo.*, Simon Fraser University.

[37] Tsay, R. (2002), *Analysis of Financial Time Series,* John Wiley & Sons, Inc.

[38] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models", *Econometrica* 50, 1-25.

[39] van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK.

[40] Zakoian, J-M. (1994), "Threshold Heteroskedasticity Models", *Journal of Economic Dynamics and Control* 18, 931-955.