

Are There Multiple Regimes in Financial Volatility?

Marcelo C. Medeiros

Department of Economics, Pontifical Catholic University of Rio de Janeiro

Alvaro Veiga

Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro

VERY PRELIMINARY AND INCOMPLETE – COMMENTS SOLICITED

November 23, 2001

Abstract

Are there multiple regimes in financial volatility? This paper addresses this question by proposing a new GARCH model with multiple regimes. The proposed formulation is an extension of the Logistic Smooth Transition GARCH model, where the conditional variance obeys different regimes with a smooth transition between them. The problems of selecting the number of regimes is solved by applying Lagrange multiplier type tests, with the aim of avoiding the estimation of unidentified models.

Keywords: Volatility, GARCH models, multiple regimes, nonlinear time series, finance.

JEL Classification Codes: C13,C22,C51,C52,C53,G15

1 Introduction

Modelling and forecasting the conditional variance, or the volatility, of financial time series is one of the major topics in financial econometrics nowadays. Forecasted conditional variances are used, for example, in portfolio selection, derivative pricing and hedging, risk management, market timing, and market making. Among solutions to tackle this problem, the ARCH (Autoregressive Conditional Heterokedasticity) model proposed by Engle (1982) and the GARCH (Generalized Autoregressive Conditional Heterokedasticity) specification introduced by Bollerslev (1986) are certainly among the most widely used and are now fully incorporated into the econometric practice. A GARCH model states that the one-step-ahead conditional variance of returns is a deterministic linear function of lagged squared values of the series and past conditional variances. For example, a GARCH(1,1) model is expressed as

$$\begin{aligned}y_t &= h_t^{1/2} \varepsilon_t, \\h_t &= \alpha + \beta h_{t-1} + \lambda y_{t-1}^2,\end{aligned}\tag{1}$$

where ε_t is usually a normal white noise with unit variance and the set of restrictions below is satisfied:

$$\alpha > 0, \beta \geq 0, \lambda \geq 0, \text{ and } \beta + \lambda < 1.$$

The linear functional form of the conditional variance in (1) is particularly convenient since it implies a linear (non-Gaussian) ARMA process for y_t^2 capturing the significant autocorrelations observed in the squared returns of financial time series, as well as the excess of kurtosis. In addition, linearity simplifies statistical inference. For good surveys on GARCH modelling, see, for example, Bera and Higgins (1993) and Bollerslev, Engle and Nelson (1994).

One drawback of the GARCH model is the symmetry in the response of volatility on past shocks, failing to accomodate sign asymmetries. Researchers, beginning with Black (1976), pointed out the asymmetric response of the conditional variance of the series to unanticipated news, represented by shocks. It is found that financial markets become more volatile in response to bad news (negative shocks) than to good news (positive shocks). Goetzmann, Ibbotson and Peng (2001) find evidence of asymmetric sign effects in volatility as far 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 equivalent positive

shocks in returns. Similar results were also reported by Schwert (1990).

There are two possible explanations for the asymmetric behavior of financial volatility: the leverage effect and the time-varying risk premium effect. In the first one it is claimed that return shocks lead to changes in the conditional volatility. When unexpected bad news arrives in the market, stock prices decline, raising the debt to equity ratio, thus raising the volatility (Black 1976, Christie 1982, Schwert 1990), causing a positive correlation between lagged negative shocks and volatility. On the other hand, in the time-varying risk premium theory, mainly motivated by the persistence in volatility and by the positive intertemporal correlation between expected return and conditional variance, it is claimed that unexpected news increase current volatility and thus risk premia and hence drive down current prices, dampening the impact of good news and amplifying the impact of bad news. This process continues until the expected return is sufficient high to compensate for the increased risk, giving rise to volatility persistence. Hence, there is a positive intertemporal correlation between unexpected changes in volatility and stock returns (Pindyck 1984, French, Schwert and Stambaugh 1987, Campbell and Hentschel 1992, Wu 1998). Which effect is the main determinant of asymmetric volatility remains an open question and will not be addressed in this paper.

As pointed out in Verhoeven and McAleer (2001), it is the average response to shocks, and not the difference in persistence of outliers and small shocks, that is modelled by GARCH models. Since the GARCH model assumes that any given realization contains information about the future evolution of volatility, estimates from the model tend to delay and then overshoot volatility increases and adjust only slowly to volatility declines. For that reason, in the case of univariate GARCH models, the standardized residuals will always have some leptokurtosis as these models are unable to capture the initial outlier. Moreover, skewness will also be affected by such characteristics.

This asymmetry has motivated a number of different ARCH/GARCH models. Among them are the EGARCH (Exponential GARCH) model proposed by Nelson (1991), the GJR-GARCH proposed by Glosten, Jagannathan and Runkle (1993), the T-GARCH (Threshold GARCH) developed by Rabemananjara and Zakoian (1993) and Zakoian (1994), and the Logistic Smooth Transition GARCH (LST-GARCH) model of Haregud (1997) and Gonzalez-Rivera (1998). Usually, only two regimes are considered: one associated with positive past returns (good news) and the other to negative past returns (bad news). The transition between regimes is either abrupt as in the T-GARCH or GJR-GARCH models or

smooth as in the LSTGARCH formulation. In any case, the switching rules between regimes can be expressed by a transition function – a step or a logistic function – centered at zero. Some papers have tried to consider more than two regimes. For example, Li and Li (1996) and Liu, Li, and Li (1997) proposed the Double Threshold GARCH (DTGARCH) model, which combines a Threshold Autoregressive (TAR) model with a T-GARCH additive error with multiple regimes. However, in their empirical illustration there is only one threshold at zero. More recently, the Tree Structured GARCH developed by Audrino and Bühlmann (in press) also considers multiple regimes but no statistical framework is adopted to test if the regimes are statistically significant.

The goal of this paper is twofold. First, we propose a general form to a multiple regime smooth transition GARCH model with no preset threshold values, in opposition to the usual zero threshold adopted by most of the models cited above. This is a more flexible data-driven generalization of the two-regime LSTGARCH model and has the main advantage to model multiple regimes in the conditional variance behavior. A similar model is discussed in Verhoeven and McAleer (2001). Second, making use of this formulation, we propose a statistical framework based on the Lagrange Multiplier (LM) test, to determine the number of regimes, very much in the spirit of Haregud (1997). Doing this we are able to check if the sign asymmetry is stronger for large shocks than for small shocks. For example, Verhoeven and McAleer (2001) show that volatility responds stronger to large negative shocks than to small negative shocks, while large positive shocks lead to a reduction in volatility. Furthermore, the size asymmetry for positive shocks is stronger than for negative shocks.

The plan of the paper is as follows. Section 2 presents the model. Section 4 describes the test for an additional regime. The modelling cycle procedure is described in Section 5. A Monte-Carlo simulation is carried out in Section 6 and an empirical example is shown in Section 7. Finally, Section 8 concludes.

2 The Model

The model proposed in this paper is called Multiple Regime Smooth Transition GARCH (MRSTGARCH) model and is defined as follows.

Definition 1 *A time series $\{y_t\}$ follows a first-order Multiple Regime Smooth Transition GARCH model,*

MRSTGARCH(1,1), if

$$\begin{aligned} y_t &= h_t^{1/2} \varepsilon_t, \\ h_t &= \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \sum_{i=1}^H \{[\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2] f(\gamma_i (s_t - c_i))\}, \end{aligned} \quad (2)$$

where $\varepsilon_t \sim NID(0, \sigma^2)$ and $f(\gamma_i (s_t - c_i))$, $i = 1, \dots, H$, is the logistic function defined as

$$f(\gamma_i (s_t - c_i)) = \frac{1}{1 + e^{-\gamma_i (s_t - c_i)}}, \quad \gamma_i > 0. \quad (3)$$

The parameter γ_i , $i = 1, \dots, H$ is called the *slope parameter*. When $\gamma_i \rightarrow \infty$, the logistic function becomes a step function. The variable s_t is known as the *transition variable*. In this paper, we consider $s_t = y_{t-1}$. Hence, we model the differences in the dynamics of the conditional volatility according to the magnitude and size of shocks in past returns.

To state the conditions to ensure strictly positive conditional variances and second-order stationarity we need to make the following assumptions:

(A.1) The parameters $c_1, \dots, c_H, \gamma_1, \dots, \gamma_H$ satisfy the conditions: $-\infty < c_1 < \dots < c_H < \infty$ and $\gamma_i > 0, i = 1, \dots, H$.

(A.2) The logistic functions satisfy the following restriction: $f(\gamma_1 (s_t - c_1)) \leq f(\gamma_2 (s_t - c_2)) \leq \dots \leq f(\gamma_H (s_t - c_H))$.

Assumption (A.1) guarantees the identifiability of the model and assumption (A.2) simplifies the conditions to ensure strictly positive variances, given in the following theorem.

Theorem 2 *Under the assumptions (A.1) and (A.2) and considering that $f(\gamma_i (s_t - c_i))$ is bounded between 0 and 1, the sufficient conditions to guarantee strictly positive conditional variance are:*

$$\mathbf{\Lambda}_1 \boldsymbol{\psi}_1 > \mathbf{0} \text{ and } \mathbf{\Lambda}_2 \boldsymbol{\psi}_2 \geq \mathbf{0} \quad (4)$$

where $\boldsymbol{\psi}_1 = [\alpha_0, \alpha_1, \dots, \alpha_H]'$, $\boldsymbol{\psi}_2 = [\beta_0, \beta_1, \dots, \beta_H, \lambda_0, \lambda_1, \dots, \lambda_H]'$, $\mathbf{\Lambda}_1$ is a lower triangular ($H +$

$1) \times (H + 1)$ matrix and Λ_2 is a $2(H + 1) \times 2(H + 1)$ block-diagonal matrix defined as

$$\Lambda_2 = \begin{pmatrix} \Lambda_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Lambda_1 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Lambda_1 \end{pmatrix}.$$

Proof. Omitted.

The conditions for a first-order MRSTGARCH process to have a finite variance and, therefore, to be second-order (covariance) stationary are given in the following theorem.

Theorem 3 *Under assumption (A.1) and the restrictions imposed in Theorem 1, the sufficient conditions for second-order stationarity of (2) is*

$$\beta_0 + \lambda_0 < 1 \quad (5)$$

and

$$\beta_0 + \lambda_0 + \sum_{i=1}^H (\beta_i + \lambda_i) < 1. \quad (6)$$

Proof. Under construction.

Although we are going to deal only with the first-order specification of model (2), it is straightforward to generalize it to a MRSTGARCH(p,q) model.

3 Parameter Estimation

As selecting the number of regimes requires estimation of MRSTGARCH models, we now turn to this problem. In this paper we estimate the parameters of our MRSTGARCH model by maximum likelihood making use of the assumptions made of ε_t in Section 2.

Using model (2) as our starting point we make the following assumption:

(A.3) The $(r \times 1)$ parameter vector

$$\boldsymbol{\psi} = [\alpha_0, \alpha_1, \dots, \alpha_H, \beta_0, \beta_1, \dots, \beta_H, \lambda_0, \lambda_1, \dots, \lambda_H, \gamma_1, \dots, \gamma_H, c_1, \dots, c_H]'$$

is an interior point of the compact parameter space Ψ which is a subspace of \mathbb{R}^r , the r -dimensional Euclidean space.

The maximum likelihood estimator of the parameters of the conditional variance equals

$$\hat{\psi} = \underset{\psi}{\operatorname{argmax}} L_T(\psi) = -\underset{\psi}{\operatorname{argmin}} \sum_{t=1}^T l_t(\psi),$$

where $l_t = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(h_t) - \frac{y_t^2}{2h_t}$. Conditions for consistency and asymptotic normality of $\hat{\psi}$ are stated in the following theorem.

Theorem 4 . *Under the assumptions (A.1)–(A.3) the maximum likelihood estimator $\hat{\psi}$ is almost surely consistent for ψ and*

$$T^{1/2}(\hat{\psi} - \psi) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}(\psi)^{-1}), \quad (7)$$

where $\mathbf{I}(\psi) = \frac{1}{T} \frac{\partial^2 L_T(\psi)}{\partial \psi \partial \psi'}$.

Proof Under construction.

In this paper, the large sample estimator of the information matrix ($\mathbf{I}(\psi)$) that we apply is

$$\hat{\mathbf{I}}(\hat{\psi}) = \frac{1}{2T} \sum_{t=1}^T \left[\frac{1}{h_t} \frac{\partial l_t(\psi)}{\partial \psi} \Big|_{\psi=\hat{\psi}} \quad \frac{\partial l_t(\psi)}{\partial \psi'} \Big|_{\psi=\hat{\psi}} \right]. \quad (8)$$

The derivatives of the log-likelihood function are shown in Appendix A.

The estimation of the MRSTGARCH model is, in general, not easy and requires the use of constrained optimization algorithms. For a description of that kind of algorithms see, for example, Bertsekas (1995).

The main difficulty in the estimation process is to obtain reasonably accurate estimates for the slope parameters. Specially when γ_i , $i = 1, \dots, H$, are very large. To obtain an accurate estimate of a large γ_i it is necessary to have a large number of observations in the neighborhood of c_i . When the sample size is not very large, there are generally few observations sufficiently close to c_i in the sample, which results in imprecise estimates of the slope parameter. This manifests itself in low absolute t -values for the estimates of γ_i . In such cases, the model builder cannot take a low absolute value of the t -statistic of the parameters of the transition function as evidence for omitting the transition function in question.

4 Determining the Number of Regimes

4.1 Testing $H = 0$ Against $H = 1$

Consider $H = 1$ and rewrite model (2) as

$$\begin{aligned} y_t &= h_t^{1/2} \varepsilon_t, \\ h_t &= \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + [\alpha_1 + \beta_1 h_{t-1} + \lambda_1 y_{t-1}^2] f(\gamma_1 (s_t - c_1)). \end{aligned} \quad (9)$$

The idea is to test the presence of a second regime represented by the term

$$[\alpha_1 + \beta_1 h_{t-1} + \lambda_1 y_{t-1}^2] f(\gamma_1 (s_t - c_1)). \quad (10)$$

A convenient null hypothesis is

$$H_0 : \gamma_1 = 0, \quad (11)$$

against the alternative $\gamma > 0$. Note that model (9) is not identified under the null hypothesis. To remedy this problem, we follow Haregud (1997) and expand the logistic function $f(\gamma_1 (s_t - c_1))$ into a first order Taylor expansion around the null hypothesis $\gamma_1 = 0$. After merging terms, the resulting model for h_t is

$$h_t = \tilde{\alpha}_0 + \tilde{\beta}_0 h_{t-1} + \tilde{\lambda}_0 y_{t-1}^2 + \pi s_t + \delta h_{t-1} s_t + \rho y_{t-1}^2 s_t, \quad (12)$$

where $\tilde{\alpha}_0 = \alpha_0 - \frac{\alpha_1 \gamma_1 c_1}{4}$, $\tilde{\beta}_0 = \beta_0 - \frac{\beta_1 \gamma_1 c_1}{4}$, $\tilde{\lambda}_0 = \lambda_0 - \frac{\lambda_1 \gamma_1 c_1}{4}$, $\pi = \frac{\gamma_1}{4}$, $\delta = \frac{\beta_1 \gamma_1}{4}$, and $\rho = \frac{\lambda_1 \gamma_1}{4}$. The null hypothesis becomes

$$H_0 : \pi = \delta = \rho = 0. \quad (13)$$

Under H_0 , $h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2$. The local approximation to the log likelihood function in a neighborhood of H_0 is

$$\begin{aligned} l_t &= -\frac{1}{2} \ln 2\pi - \frac{1}{2} \left(\tilde{\alpha}_0 + \tilde{\beta}_0 h_{t-1} + \tilde{\lambda}_0 y_{t-1}^2 + \pi s_t + \delta h_{t-1} s_t + \rho y_{t-1}^2 s_t \right) - \\ &\quad \frac{y_t^2}{\left(\tilde{\alpha}_0 + \tilde{\beta}_0 h_{t-1} + \tilde{\lambda}_0 y_{t-1}^2 + \pi s_t + \delta h_{t-1} s_t + \rho y_{t-1}^2 s_t \right)}. \end{aligned} \quad (14)$$

Theorem 5 Under H_0 , the LM statistic is

$$LM = \frac{1}{2} \left\{ \sum_{t=1}^T \frac{1}{\hat{h}_{0,t}} \left(\frac{y_t^2}{\hat{h}_{0,t}} - 1 \right) \hat{\mathbf{z}}_t \right\} \times \left\{ \sum_{t=1}^T \frac{1}{\hat{h}_{0,t}^2} \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t' \right\}^{-1} \times \left\{ \sum_{t=1}^T \frac{1}{\hat{h}_{0,t}} \left(\frac{y_t^2}{\hat{h}_{0,t}} - 1 \right) \hat{\mathbf{z}}_t \right\}, \quad (15)$$

where $\hat{h}_{0,t} = \hat{\alpha}_0 + \hat{\beta}_0 \hat{h}_{0,t-1} + \hat{\lambda}_0 y_{t-1}^2$ is the estimated variance under the null of a GARCH(1,1) model, $\hat{\mathbf{z}}_t = \sum_{i=1}^t \tilde{\beta}_0^{t-i} \hat{\mathbf{x}}_i$, and $\hat{\mathbf{x}}_t = [1, \hat{h}_{0,t-1}, y_{t-1}^2, s_t, \hat{h}_{0,t-1} s_t, y_{t-1}^2 s_t]$, has a χ^2 distribution with 3 degrees of freedom.

Proof. See Appendix B.

The test can be carried out in stages as follows.

1. Estimate a GARCH(1,1) model for y_t , call the estimated variance $\hat{h}_{0,t}$. At this point, a problem concerning the estimation of the model under the null must be considered. Given that the parameters of the model are estimated by maximum likelihood, the standardized residuals $\frac{y_t^2}{\hat{h}_{0,t}}$ should be orthogonal to

$$\frac{\hat{\mathbf{z}}_t^*}{\hat{h}_{0,t}} = \frac{\sum_{i=1}^t \tilde{\beta}_0^{t-i} \hat{\mathbf{x}}_i^*}{\hat{h}_{0,t}},$$

where $\hat{\mathbf{x}}_i^* = [1, \hat{h}_{0,i-1}, y_{i-1}^2]'$. This should be independent of whether the null hypothesis hold or not. However, in practical situations, exact orthogonality is not always guaranteed and the size of the test may be distorted. To overcome this problem, the standard procedure is to first regress $\left(\frac{y_t^2}{\hat{h}_{0,t}} - 1 \right)$ on $\frac{\hat{\mathbf{z}}_t^*}{\hat{h}_{0,t}}$ and then compute $SSR_0 = \sum_{t=1}^T \hat{u}_t$, where \hat{u}_t is the estimated residuals from the regression.

2. Compute $SSR_1 = \sum_{t=1}^T \hat{v}_t$, where \hat{v}_t is the estimated residuals from the regression of \hat{u}_t on $\frac{\hat{\mathbf{z}}_t}{\hat{h}_{0,t}}$.
3. Compute the χ^2 statistic

$$LM_{\chi^2}^H = T \frac{SSR_0 - SSR_1}{SSR_0}, \quad (16)$$

or the F statistic

$$LM_F^H = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - m - n)}, \quad (17)$$

where $n = 3$ and $m = 3$ are respectively the number of regressors under the null and alternative hypothesis.

Under H_0 , $LM_{\chi^2}^H$ is approximately distributed as a χ^2 with m degrees of freedom and LM_F^H has approximately an F distribution with m and $T - m - n$ degrees of freedom.

4.2 Testing for $H > 1$

Consider the following model

$$\begin{aligned}
y_t &= h_t^{1/2} \varepsilon_t, \\
h_t &= \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \\
&\quad \sum_{i=1}^H \{ [\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2] f(\gamma_i (s_t - c_i)) \} + \\
&\quad [\alpha_{H+1} + \beta_{H+1} h_{t-1} + \lambda_{H+1} y_{t-1}^2] f(\gamma_{H+1} (s_t - c_{H+1})).
\end{aligned} \tag{18}$$

In order to test the additional regime, a convenient null hypothesis is

$$\gamma_{H+1} = 0, \tag{19}$$

against the alternative $\gamma_{H+1} > 0$. Again model (18) is not identified under the null hypothesis. To circumvent the problem, we proceed in the same fashion as before, expanding the additional transition function $f(\gamma_{H+1} (s_t - c_{H+1}))$ into a first order Taylor expansion around the null hypothesis $\gamma_{H+1} = 0$. After merging terms, the resulting model for h_t is

$$\begin{aligned}
h_t &= \tilde{\alpha}_0 + \tilde{\beta}_0 h_{t-1} + \tilde{\lambda}_0 y_{t-1}^2 + \sum_{i=1}^H \{ [\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2] f(\gamma_i (s_t - c_i)) \} + \\
&\quad \pi s_t + \delta h_{t-1} s_t + \rho y_{t-1}^2 s_t,
\end{aligned} \tag{20}$$

where $\tilde{\alpha}_0 = \alpha_0 - \frac{\alpha_{H+1} \gamma_{H+1} c_{H+1}}{4}$, $\tilde{\beta}_0 = \beta_0 - \frac{\beta_{H+1} \gamma_{H+1} c_{H+1}}{4}$, $\tilde{\lambda}_0 = \lambda_0 - \frac{\lambda_{H+1} \gamma_{H+1} c_{H+1}}{4}$, $\pi = \frac{\gamma_{H+1}}{4}$, $\delta = \frac{\beta_{H+1} \gamma_{H+1}}{4}$, and $\rho = \frac{\lambda_{H+1} \gamma_{H+1}}{4}$.

The null hypothesis becomes

$$H_0 : \pi = \delta = \rho = 0. \tag{21}$$

The LM statistic is (15) with

$$\hat{\mathbf{z}}_t = \hat{\mathbf{x}}_t + \sum_{k=1}^{t-1} \left\{ \prod_{j=k+1}^t \left[\hat{\beta}_0 + \sum_{i=1}^H \hat{\beta}_i f(\hat{\gamma}_j(s_j - \hat{c}_i)) \right] \right\} \hat{\mathbf{x}}_k,$$

where $\hat{\mathbf{x}}_t$ is the estimated gradient of equation (20) and $\hat{h}_{0,t}$ is replaced by

$$\hat{h}_{H,t} = \hat{\alpha}_0 + \hat{\beta}_0 \hat{h}_{H,t-1} + \hat{\lambda}_0 y_{t-1}^2 + \sum_{i=1}^H \left\{ \left[\hat{\alpha}_i + \hat{\beta}_i \hat{h}_{H,t-1} + \hat{\lambda}_i y_{t-1}^2 \right] f(\hat{\gamma}_i(s_t - \hat{c}_i)) \right\}.$$

The test is carried out in stages as before. The only differences are the new definitions of \mathbf{z}_t and \mathbf{x}_t , the replacement of $\hat{h}_{0,t}$ by $\hat{h}_{H,t}$, and the new degrees of freedom of the test distributions, $m = 3$ and $n = 3 + 2H$.

When applying the test a special care should be taken. If γ_i is large, we may have numerical problems when carrying out the test in small samples. A solution is to omit, in those cases, the terms that depend on the derivatives of the logistic function from the test statistic. This can be done without significantly affecting the value of the test statistic.

5 Modelling Cycle

At this point we are ready to combine the above statistical ingredients into a coherent modelling strategy. We begin testing linearity against an ARCH(q) model ¹ at significance level α . The model under the null hypothesis is simply an white noise. If the null hypothesis is not rejected, the white noise is accepted. In case of a rejection, a GARCH(1,1) model is estimated and tested against a model with MRSTGARCH(1,1) with two regimes at the significance level $\alpha\varrho$, $0 < \varrho < 1$. Another rejection leads to estimating a model with two regimes and testing it against a model with three three at the significance level $\alpha\varrho^2$. The sequence is terminated at the first acceptance of the null hypothesis. The significance level is reduced at each step of the sequence and converges to zero. This way we avoid excessively large models and control the overall significance level of the procedure. An upper bound for the overall significance level α^* may be obtained using the Bonferroni bound. For example, if $\alpha = 0.1$, and $\varrho = 1/2$

¹Bollerslev (1986) pointed out that under the null of no heteroskedasticity, there is no general Lagrange Multiplier test for GARCH(p,q). This is due to the fact that the Hessian is singular if both $p > 0$ and $q > 0$.

then $\alpha^* \leq 0.187$. Note that if we instead of our LM type test apply a model selection criterion such as AIC or BIC to this sequence, we in fact use the same significance level at each step.

6 Monte-Carlo Experiment

In this section we report the results from a Monte-Carlo experiment for testing the empirical size and power of the test (F-version) described in Section 4.1 to detect a second regime in the conditional variance of the data. The size simulations are based on data with 500 observations generated from different GARCH(1,1) specifications (models A–D below). To investigate the power of the test we simulate data generated from a MRSTGARCH(1,1) process (model E below). The results are evaluated over 1000 replications of each model. The used models are:

1. Model A:

$$\text{GARCH}(1,1): \alpha = 0.5 \times 10^{-6}, \beta = 0.70, \lambda = 0.25.$$

2. Model B:

$$\text{GARCH}(1,1): \alpha = 1.0 \times 10^{-5}, \beta = 0.85, \lambda = 0.05.$$

3. Model C:

$$\text{GARCH}(1,1): \alpha = 1.0 \times 10^{-5}, \beta = 0.90, \lambda = 0.05.$$

4. Model D:

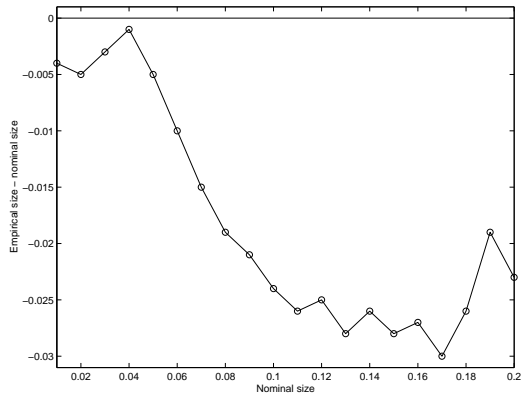
$$\text{GARCH}(1,1): \alpha = 1.0 \times 10^{-5}, \beta = 0.90, \lambda = 0.09.$$

5. Model E:

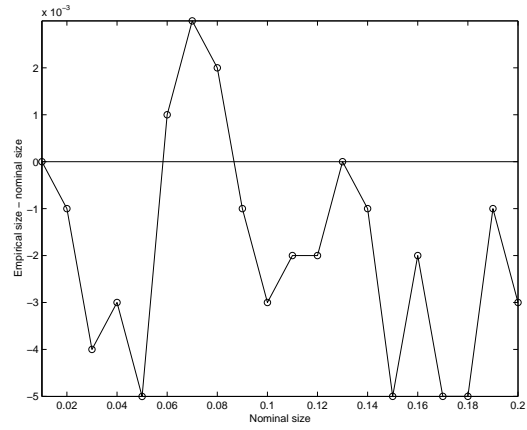
$$\begin{aligned} \text{MRSTGARCH}(1,1): \alpha_0 = 5.0 \times 10^{-6}, \beta_0 = 0.85, \lambda_0 = 0.01, \alpha_1 = 1.0 \times 10^{-5}, \beta_1 = 0.05, \\ \lambda_1 = 0.09, \gamma_1 = 300, c_1 = 0. \end{aligned}$$

Figure 1 shows the size-discrepancy plot for models A–D. The empirical size of the test is relatively close to the nominal size of the test.

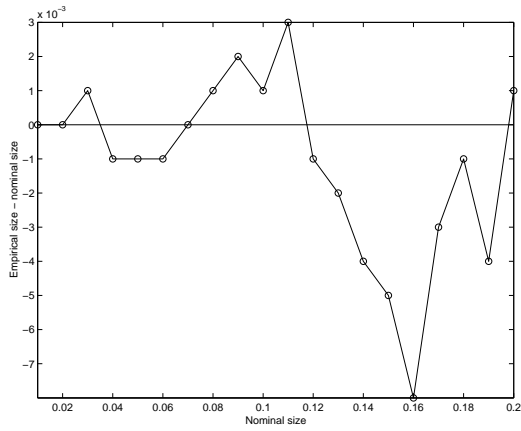
The power-size curve is shown in Figure 2. As can be seen by inspection of Figure 2, the test has good power properties.



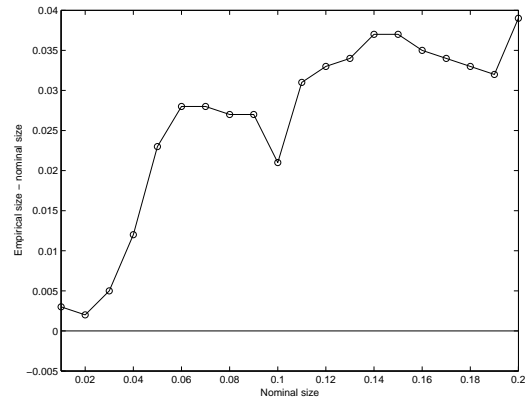
(a)



(b)



(c)



(d)

Figure 1: Size-discrepancy plots. Panel (a) refers to model A. Panel (b) refers to model B. Panel (c) refers to model C. Panel (d) refers to model D.

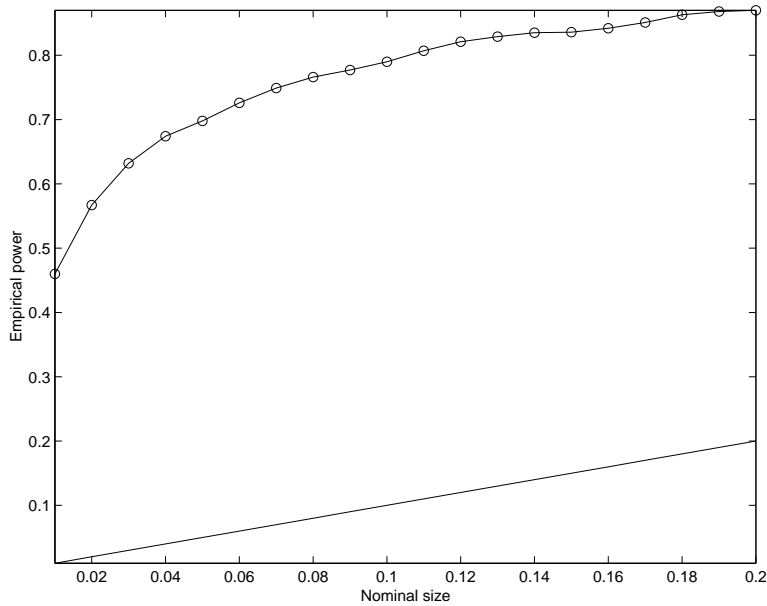


Figure 2: Power-size curve.

7 Empirical Examples

The data that are used in this paper consist of eight indices of major stock markets. We employ the indices of the stock markets in Amsterdam (EOE), Frankfurt (DAX), Hong Kong (Hang Seng), London (FTSE100), New York, (S&P 500), Paris (CAC40), Singapore (Singapore All Shares) and Tokyo (Nikkei). The data are obtained from Dick van Dijk’s home page ² and is the same used in Franses and van Dijk (2000). Table 1 lists the series, the sample period, and the number of observations. Descriptive statistics for each series are shown in Table 2.

We start estimating a GARCH(1,1) model for each series. The results are shown in Table 3. Table 3 shows, respectively, four goodness-of-fit statistics, the p -value of the Jarque-Bera test on the standardized residuals, and the p -value of the F-version of the LM test for the second regime as described in Section 4.1 and a reduced version of the test where only the squared return coefficient changes according to the regime ($\alpha_i = \beta_i = 0, i = 1, \dots, h$). In that case, the test is equivalent to the one proposed in Haregud (1997). The goodness-of-fit statistics shown in Table 3 are:

- The logarithm of the normal likelihood. The best predictor is considered the one with the highest

²<http://www.few.eur.nl/few/people/djvandijk/nltsmef/nltsmef.htm>

Table 1: Data sets.

Series	Stock Market	Period	Number of Observations
1 - EOE	Amsterdam	01/07/1986 – 12/31/1997	3127
2 - DAX	Frankfurt	01/07/1986 – 12/31/1997	3127
3 - Hang Seng	Hong Kong	01/07/1986 – 12/31/1997	3127
4 - FTSE100	London	01/07/1986 – 12/31/1997	3127
5 - S&P 500	New York	01/07/1986 – 12/31/1997	3127
6 - CAC40	Paris	07/09/1987 – 12/31/1997	2735
7 - Singapore All Shares	Singapore	01/07/1986 – 12/31/1997	3127
8 - Nikkei	Tokyo	01/07/1986 – 12/31/1997	3127
9 - IBOVESPA	São Paulo	01/02/1995 – 10/11/2001	1678

Table 2: Descriptive statistics.

Series	Mean	Standard Deviation	Asymmetry	Kurtosis
1 - EOE	3.83×10^{-4}	0.01	-0.69	19.80
2 - DAX	3.49×10^{-4}	0.01	-0.95	15.07
3 - Hang Seng	5.71×10^{-4}	0.02	-5.00	119.28
4 - FTSE100	4.10×10^{-4}	0.01	-1.59	27.42
5 - S&P 500	4.89×10^{-4}	0.01	-4.30	99.71
6 - CAC40	2.58×10^{-4}	0.01	-0.53	10.56
7 - Singapore All Shares	1.92×10^{-4}	0.01	-0.25	28.16
8 - Nikkei	4.99×10^{-5}	0.01	-0.21	14.80
9 - IBOVESPA	4.46×10^{-4}	0.03	0.70	15.80

value of the log-likelihood in the estimation period.

$$LKHD = -\frac{1}{2T} \sum_{t=1}^T \left[\frac{y_t^2}{\hat{h}_t} + \ln(\hat{h}_t) \right]. \quad (22)$$

- The root mean squared error (RMSE) of the square of the observations. The best predictor is considered the one with the lowest RMSE of the squared observations.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t^2 - \hat{h}_t)^2}. \quad (23)$$

As pointed out in Lopez (2001) and Bollerslev et al. (1994) the RMSE has two important shortcomings. First, the use of y_t^2 proxy is imprecise and noisy due to its asymmetric distribution. Second the symmetric nature of this loss function does not sufficiently penalize non-positive variance forecasts. To circumvent those difficulties, we also use two asymmetric loss functions.

- The heteroskedastic-adjusted mean squared error (HMSE) (Bollerslev and Ghysels 1996)

$$HMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\frac{y_t^2}{\hat{h}_t} - 1 \right)^2}. \quad (24)$$

- The Logarithmic Loss (LL) (Pagan and Scwert 1990)

$$LL = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\log(y_t^2) - \log(\hat{h}_t) \right)^2}. \quad (25)$$

We reject the null of a GARCH(1,1) model against a MRSTGARCH(1,1) for six out of the nine original series. We proceed estimating the MRSTGARCH(1,1) model for each one of the six series. The results are shown in Table 4.

8 Conclusions

In this paper we put forward a new nonlinear GARCH(1,1) model to describe the asymmetric behavior observed in financial time series, specially, stock returns. The model is called the Multiple Regime

Table 3: Results of the GARCH estimation.

Series	LKHD	RMSE	LL	HMSE	JB	LM (Full)	LM (Reduced)
1 - EOE	-4.16	7.90×10^{-4}	8.97	11.79	0	5.00×10^{-3}	0.08
2 - DAX	-4.01	9.39×10^{-4}	8.79	17.48	0	0.07	0.18
3 - Hang Seng	-3.83	3.00×10^{-2}	9.59	16.77	0	5.00×10^{-3}	4.00×10^{-3}
4 - FTSE100	-4.29	5.10×10^{-4}	7.22	16.19	0	0.12	0.19
5 - S&P 500	-4.29	3.00×10^{-3}	9.00	10.96	0	1.96×10^{-4}	0.59
6 - CAC40	-4.02	4.77×10^{-4}	9.74	5.60	0	8.47×10^{-5}	9.46×10^{-4}
7 - Singapore All Shares	-4.20	9.10×10^{-4}	10.84	32.93	0	0.20	0.09
8 - Nikkei	-3.95	1.40×10^{-3}	13.49	11.81	0	5.45×10^{-5}	3.00×10^{-3}
9 - IBOVESPA	-3.30	1.41×10^{-2}	6.48	3.35	0	1.79×10^{-11}	3.31×10^{-9}

Table 4: Results of the MRSTGARCH estimation.

Series	LKHD	RMSE	LL	HMSE	JB	Number of Regimes
1 - EOE	-4.18	7.72×10^{-4}	8.85	6.78	0	4
2 - Hang Seng	-3.85	4.54×10^{-2}	9.44	9.00	0	2
3 - S&P 500	-4.29	3.00×10^{-3}	8.88	7.67	0	4
4 - CAC40	-4.02	4.63×10^{-4}	9.74	4.91	0	2
5 - Nikkei	-3.99	1.30×10^{-3}	13.18	6.70	0	6
6 - IBOVESPA	-3.32	1.26×10^{-2}	6.23	3.27	0	4

Smooth Transition GARCH (MRSTGARCH) model and is a straightforward generalization of the Logistic Smooth Transition GARCH (LSTGARCH) model, being capable of modelling multiple regimes in the conditional variance of the series. A coherent sequence of Lagrange Multiplier (LM) tests were developed to determine the number of regimes. A simulation experiment was carried out to check the size and power properties of the proposed tests. An empirical illustration with nine stock indexes were carried out to evaluate the performance of the proposed model in comparison with the standard GARCH(1,1) specification. We have found strong evidence of asymmetry in six out of nine series.

A Derivatives of the Log-Likelihood Function

As ε_t in model (2) is normally distributed, l_t is given by

$$l_t = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(h_t) - \frac{y_t^2}{2h_t}. \quad (26)$$

Taking the first derivatives with respect to the parameters we get

$$\frac{\partial l_t}{\partial \boldsymbol{\psi}} = \frac{1}{2h_t} \left(\frac{y_t^2}{h_t} - 1 \right) \frac{\partial h_t}{\partial \boldsymbol{\psi}}. \quad (27)$$

Defining $G_\beta(s_t; \boldsymbol{\psi}) = \beta_0 + \sum_{i=1}^h \beta_i f(\gamma_i(s_t - c_i))$ and conditioned on $\frac{\partial h_0}{\partial \boldsymbol{\psi}} = \mathbf{0}$ the partial derivatives of h_t with respect to the parameters are given by:

$$\frac{\partial h_t}{\partial \alpha_0} = 1 + \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] \quad (28)$$

$$\frac{\partial h_t}{\partial \beta_0} = h_{t-1} + \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] h_{t-k} \quad (29)$$

$$\frac{\partial h_t}{\partial \lambda_0} = y_{t-1}^2 + \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] y_{t-k}^2 \quad (30)$$

$$\frac{\partial h_t}{\partial \alpha_i} = f(\gamma_i(s_t - c_i)) + \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] f(\gamma_i(s_k - c_i)), \quad i = 1, \dots, H \quad (31)$$

$$\frac{\partial h_t}{\partial \beta_i} = f(\gamma_i(s_t - c_i))h_{t-1} + \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] f(\gamma_i(s_k - c_i))h_{t-k}, \quad i = 1, \dots, H \quad (32)$$

$$\frac{\partial h_t}{\partial \lambda_i} = f(\gamma_i(s_t - c_i))y_{t-1}^2 + \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] f(\gamma_i(s_k - c_i))y_{t-k}^2, \quad i = 1, \dots, H \quad (33)$$

$$\begin{aligned} \frac{\partial h_t}{\partial \gamma_i} &= (\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2) f(\gamma_i(s_t - c_i)) [1 - f(\gamma_i(s_t - c_i))] (s_t - c_i) + \\ &\quad \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] (\alpha_i + \beta_i h_{t-k} + \lambda_i y_{t-k}^2) \times \\ &\quad f(\gamma_i(s_k - c_i)) [1 - f(\gamma_i(s_k - c_i))] (s_k - c_i), \quad i = 1, \dots, H \end{aligned} \quad (34)$$

$$\begin{aligned} \frac{\partial h_t}{\partial c_i} = & -(\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2) f(\gamma_i(s_t - c_i)) [1 - f(\gamma_i(s_t - c_i))] - \\ & \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^t G_\beta(s_j; \boldsymbol{\psi}) \right] (\alpha_i + \beta_i h_{t-k} + \lambda_i y_{t-k}^2) \times \\ & f(\gamma_i(s_k - c_i)) [1 - f(\gamma_i(s_k - c_i))], \quad i = 1, \dots, H \end{aligned} \quad (35)$$

B Proof of Theorem 3

Rewrite equation (14) as

$$l_t = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \boldsymbol{\psi}' \tilde{\mathbf{x}}_t - \frac{y_t^2}{2\boldsymbol{\psi}' \tilde{\mathbf{x}}_t}, \quad (36)$$

where \mathbf{x}_t is defined as in (15) and $\boldsymbol{\psi}$ is a parameter vector defined as $\boldsymbol{\psi} = [\tilde{\alpha}_0, \tilde{\beta}_0, \tilde{\lambda}_0, \pi, \delta, \rho]'$.

The LM statistic has the general form

$$LM = T \bar{\mathbf{q}}_T(\boldsymbol{\psi})' |_{H_0} \mathbf{I}(\boldsymbol{\psi})^{-1} |_{H_0} \bar{\mathbf{q}}_T(\boldsymbol{\psi}) |_{H_0}, \quad (37)$$

where $\bar{\mathbf{q}}_T(\boldsymbol{\psi})$ is the average score and $\mathbf{I}(\boldsymbol{\psi})$ is the information matrix.

It is straightforward to show that

$$\bar{\mathbf{q}}_T(\boldsymbol{\varrho}) = \frac{1}{T} \sum_{t=1}^T \frac{1}{2h_t} \left(\frac{y_t^2}{h_t^2} - 1 \right) \frac{\partial h_t}{\partial \boldsymbol{\psi}}. \quad (38)$$

The population information matrix is defined as the negative expectation of the average Hessian.

$$\mathbf{I}(\boldsymbol{\varrho}) = -E \left(\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right), \quad (39)$$

where

$$\frac{\partial^2 l_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} = \frac{1}{2h_t^2} \left(1 - \frac{2y_t^2}{h_t} \right) \left(\frac{\partial h_t}{\partial \boldsymbol{\psi}} \right) \left(\frac{\partial h_t}{\partial \boldsymbol{\psi}} \right)' + \frac{1}{2h_t} \left(\frac{y_t^2}{h_t} - 1 \right) \frac{\partial^2 h_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}. \quad (40)$$

Combining (39) with (40), the population information matrix becomes

$$\mathbf{I}(\boldsymbol{\varrho}) = \frac{1}{2T} E \left(\sum_{t=1}^T \frac{1}{h_t^2} \left(\frac{\partial h_t}{\partial \boldsymbol{\psi}} \right) \left(\frac{\partial h_t}{\partial \boldsymbol{\psi}} \right)' \right), \quad (41)$$

where

$$\frac{\partial h_t}{\partial \boldsymbol{\psi}} = \mathbf{x}_t + \sum_{i=1}^{t-1} \left\{ \left[\prod_{j=i+1}^t (\tilde{\beta}_0 + \delta s_j) \right] \mathbf{x}_i \right\}, \quad (42)$$

conditioned on $\frac{\partial h_0}{\partial \boldsymbol{\psi}} = \mathbf{0}$.

Under the null, $\tilde{\beta}_0 = \beta_0$ and the average score vector and the population information matrix can be consistently estimated as

$$\hat{\mathbf{q}}_T(\boldsymbol{\psi})|_{H_0} = \frac{1}{T} \sum_{t=1}^T \left[\frac{1}{\hat{h}_{0,t}} \left(\frac{y_t^2}{\hat{h}_{0,t}} - 1 \right) \left(\sum_{i=1}^t \hat{\beta}_0^{t-i} \hat{\mathbf{x}}_i \right) \right] \quad (43)$$

and

$$\hat{\mathbf{I}}(\boldsymbol{\psi})|_{H_0} = \frac{1}{2T} \sum_{t=1}^T \left[\frac{1}{\hat{h}_{0,t}} \left(\sum_{i=1}^t \hat{\beta}_0^{t-i} \hat{\mathbf{x}}_i \right) \left(\sum_{i=1}^t \hat{\beta}_0^{t-i} \hat{\mathbf{x}}_i \right)' \right]. \quad (44)$$

The LM statistic can therefore be written as

$$\begin{aligned} LM = & \frac{1}{2} \left\{ \sum_{t=1}^T \left[\frac{1}{\hat{h}_{0,t}} \left(\frac{y_t^2}{\hat{h}_{0,t}} - 1 \right) \left(\sum_{i=1}^t \hat{\beta}_0^{t-i} \hat{\mathbf{x}}_i \right) \right] \right\}' \times \\ & \left\{ \sum_{t=1}^T \left[\frac{1}{\hat{h}_{0,t}} \left(\sum_{i=1}^t \hat{\beta}_0^{t-i} \hat{\mathbf{x}}_i \right) \left(\sum_{i=1}^t \hat{\beta}_0^{t-i} \hat{\mathbf{x}}_i \right)' \right] \right\}^{-1} \times \\ & \left\{ \sum_{t=1}^T \left[\frac{1}{\hat{h}_{0,t}} \left(\frac{y_t^2}{\hat{h}_{0,t}} - 1 \right) \left(\sum_{i=1}^t \hat{\beta}_0^{t-i} \hat{\mathbf{x}}_i \right) \right] \right\} \end{aligned} \quad (45)$$

(Q.E.D)

References

- Audrino, F. and Bühlmann, P. (in press). Tree-structured GARCH models, *Journal of the Royal Statistical Society, Series B*.
- Bera, A. and Higgins, M. L. (1993). ARCH models: Properties, estimation and testing, *Journal of Economic Surveys* **7**: 305–366.
- Bertsekas, D. P. (1995). *Nonlinear Programming*, Athena Scientific, Belmont, MA.

- Black, F. (1976). Studies in stock price volatility changes, *Proceedings of The American Statistical Association, Business and Economic Statistics Section* pp. 177–181.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **21**: 307–328.
- Bollerslev, T. and Ghysels, E. (1996). Periodic autoregressive conditional heteroskedasticity, *Journal of Business and Economic Statistics* **14**: 139–157.
- Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994). ARCH models, in R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier Science, pp. 2959–3038.
- Campbell, J. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* **31**: 281–318.
- Christie, A. A. (1982). The stochastic behavior of common stock variances – value, leverage and interest rates effects, *Journal of Financial Economics* **10**: 407–432.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflations, *Econometrica* **50**: 987–1007.
- Franses, P. H. and van Dijk, D. (2000). *Nonlinear Time Series Models in Empirical Finance*, Cambridge University Press.
- French, K., Schwert, G. W. and Stambaugh, R. (1987). Expected stock returns and volatility, *Journal of Financial Economics* **19**: 3–29.
- Glosten, L., Jagannathan, R. and Runkle, R. (1993). On the relationship between the expected value and the volatility of the nominal excess returns on stocks, *Journal of Finance* **48**: 1779–1801.
- Goetzmann, W. N., Ibbotson, R. G. and Peng, L. (2001). A new historical database for the nyse 1815 to 1925: Performance and predictability, *Journal of Financial Markets* **4**: 1–32.
- Gonzalez-Rivera, G. (1998). Smooth transition GARCH models, *Studies in Nonlinear Dynamics and Econometrics* **3**(2): 61–78.
- Haregud, G. E. (1997). *A New Non-Linear GARCH Model*, Phd. thesis, Stockholm School of Economics.

- Li, C. W. and Li, W. K. (1996). On a double threshold autoregressive heteroscedastic time series model, *Journal of Applied Econometrics* **11**: 253–274.
- Liu, J., Li, W. K., and Li, C. W. (1997). On a threshold autoregression with conditional heteroscedastic variances, *Journal of Statistical Planning and Inference* **62**(2): 279–300.
- Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models, *Journal of Forecasting* **20**: 87–109.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**: 347–370.
- Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatility, *Journal of Econometrics* **45**: 267–290.
- Pindyck, R. S. (1984). Risk, inflation, and the stock market, *American Economic Review* **74**: 334–351.
- Rabemananjara, R. and Zakoian, J. M. (1993). Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics* **8**(1): 31–49.
- Schwert, G. W. (1990). Stock volatility and the crash of '87, *Review of Financial Studies* **3**: 77–102.
- Verhoeven, P. and McAleer, M. (2001). Modeling conditional volatility using flexible smooth transition garch processes, *Working paper*, Department of Economics, University of Western Australia.
- Wu, G. (1998). The determinants of asymmetric volatility, *Working paper*, Graduate School of Business, Stanford University.
- Zakoian, J. M. (1994). Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* **18**: 931–955.