

CORE DISCUSSION PAPER

2006/11

## REGIME SWITCHING GARCH MODELS

Luc Bauwens<sup>1</sup>, Arie Preminger,<sup>2</sup> and Jeroen V.K. Rombouts<sup>3</sup>

June 2005, corrected July 17, 2006

### Abstract

We develop univariate regime-switching GARCH (RS-GARCH) models wherein the conditional variance switches in time from one GARCH process to another. The switching is governed by a time-varying probability, specified as a function of past information. We provide sufficient conditions for geometric ergodicity and existence of moments. Because of path dependence, maximum likelihood estimation is not feasible. By enlarging the parameter space to include the state variables, Bayesian estimation using a Gibbs sampling algorithm is feasible. We apply this model using the NASDAQ daily return series.

**Keywords:** GARCH, regime-switching, Bayesian inference.

**JEL Classification:** C11, C22, C52.

---

<sup>1</sup>CORE and Department of Economics, Université catholique de Louvain.

<sup>2</sup>CORE, Université catholique de Louvain.

<sup>3</sup>HEC Montréal and CIRANO, 3000 Cote Sainte Catherine, H3T 2A7 Montréal (QC), Canada.

The authors thank Richard Baillie, Eric Renault, Sharon Rubin, and the participants of the International Conference on High Frequency Finance in Konstanz 2006, for their helpful comments. Arie Preminger acknowledges the financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, Microstructure of Financial Markets in Europe, and the Ernst Foundation. This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

# 1 Introduction

Over the past two decades there has been a large amount of theoretical and empirical research on modelling volatility in financial markets. Since volatility is commonly used as a measure of risk associated with financial returns, it is important to portfolio managers, option traders and market makers among others. Further, portfolio optimization, derivative pricing and risk management, such as Value-at-Risk (VaR), use volatility estimates as inputs. So far in the literature, the most widespread approach to modeling volatility consists of the GARCH model of Bollerslev (1986) and its numerous extensions that can account for the volatility clustering and excess kurtosis found in the data (see e.g. Bollerslev and Wooldridge (1992) for an overview of the GARCH literature).

The accumulated evidence from empirical research suggests that the volatility of financial markets displays some type of persistence that cannot be appropriately captured by classical GARCH models. In particular, these models usually indicate high persistence in the conditional volatility. This persistence, as was noted by Hamilton and Susmel (1994), Gray (1996), and Klaassen (2002), is not compatible with the poor forecasting results of these models. Furthermore, Diebold (1986) and Lamoureux and Lastrapes (1990), among others, argue that the near integrated behavior of the conditional variance may originate from structural changes in the variance process, which are not accounted for by standard GARCH models. Mikosch and Starica (2004) show that estimating a GARCH(1,1) model on a sample displaying structural changes in the unconditional volatility does indeed create an integrated GARCH (IGARCH) effect. These findings clearly indicate a potential source of misspecification, to the extent that the structural form of the conditional mean and variance is relatively inflexible and held fixed throughout the entire sample period. For example, the existence of shifts in the variance process over time can induce volatility persistence (see Wong and Li (2001) and Lanne and Saikkonen (2003)). Hence the estimates of a GARCH model suffer from a substantial upward bias in the persistence parameter. Therefore, models in which the parameters are allowed to change over time may be more appropriate for modelling volatility.

In this perspective, several models that are based on a mixture of distributions have been proposed. Schwert (1989) considers a model in which returns may have either a high or a low variance, and switches between these states are determined by a two-state Markov process. Hamilton and Susmel (1994) and Cai (1994) introduce an ARCH model with regime-switching

parameters in order to take into account sudden changes in volatility. They use an ARCH specification instead of a GARCH to avoid the problem of path dependence of the conditional volatility on the ruling regime. Later, a tractable Markov-switching GARCH model was presented by Gray (1996) and a modification of his model was suggested by Klaassen (2002), see also Bollen, Gray, and Whaley (2000), Dueker (1997) and Haas, Mittnik, and Paolella (2004b). Several authors have also examined the class of mixtures of normal GARCH models, i.e. models where errors have a conditional distribution that is a mixture of normal ones with GARCH variance components and the probability that each observation belongs to a given volatility regime is constant. Vlaar and Palm (1993) were the first to suggest a mixture of two normal distributions where the difference between the conditional variances in each state is constant. Another version was proposed by Bauwens, Bos, and van Dijk (1999) who consider a mixture GARCH in which the two conditional variances are proportional to each other. Recently, Haas, Mittnik, and Paolella (2004a) specified a general framework for these models, allowing for interdependence between the variance components in each regime.

The objective of this paper is to develop models that better describe the volatility behavior and to extend the recent literature on switching volatility models. We propose a regime-switching GARCH (RS-GARCH) model, in which the parameters are different in each regime to account for the possibility that the data generating process undergoes a finite number of changes over the sample period. This model allows for a different speed of reversion to a different volatility level at different times throughout the sample period. The probability that each observation belongs to a given volatility regime is allowed to change over time according to the past history of the process. We provide sufficient conditions for geometric ergodicity and existence of moments of the proposed model.

The RS-GARCH model allows for periods in which the process is characterized by integrated volatility or even by non-stationarity, giving the impression of structural instability in the conditional volatility. The probabilities of non-stable regimes can be quite high in tranquil periods which can explain the substantial growth in volatility in a short amount of time. However, the process endogenously collapses back from its explosive component towards a stable regime. This occurs because the probability of staying in a non-stable regime tends to zero as the level of volatility increases, which forces the process to be globally stationary. That is, large shocks have the effect of ‘relieving pressure’ by reducing the probability of a large

shock in the next period, so that large shocks are less persistent than moderate shocks. This type of behavior is in line with a well documented feature of volatility in financial markets, namely, the occurrence of sharp upward jumps in volatility, followed by fairly rapid reversion to near normal levels of volatility, see e.g. Dueker (1997).

The computation of the likelihood function of a RS-GARCH model is not feasible due to the well-known problem of path dependence. The path dependence occurs because the conditional variance at time  $t$  depends on the entire sequence of past regimes (or states) up to time  $t$ , due to the recursive nature of the GARCH process. Since the regimes are unobservable, we need to integrate over all possible paths when computing the sample likelihood. But the number of possible regime paths grows exponentially with time, which renders the computation of the likelihood function is not feasible. To circumvent the path dependence problem, a variety of alternative tractable models that try to preserve the nature of the GARCH process have been proposed. For example, Gray (1996) replaces the lagged conditional variance by its conditional expectation with respect to the unobserved state variable. Dueker (1997) applies a collapsing procedure based on Kim's algorithm (Kim (1994)), but he essentially adopts the same solution as Gray. Klaassen (2002) modifies Gray's approach by taking the conditional expectation with respect to a broader information set than Gray (1996). Haas, Mittnik, and Paoletta (2004a and 2004b) assume that the current value of the conditional variance in each regime depends directly on the lagged value of the same conditional variance in the previous regime, i.e. they assume the parallel existence of several conditional variances. Our solution to the path dependence problem is to use Bayesian inference for the RS-GARCH model. The unobservable state variables are treated as part of the model parameters, which makes the construction of the likelihood function feasible. We construct a Gibbs sampler algorithm to compute the posterior densities. Therefore, our solution is rigorous in the sense that we estimate the RS-GARCH model exactly as specified, without resorting to Gray's or Dueker's approximations.

In Section 2, we define the RS-GARCH model and we state sufficient conditions for stochastic stability and existence of moments. In Section 3, we explain how the model can be estimated in the Bayesian framework and provide a numerical example based on simulated data. In Section 4, we apply our approach to a long time series of returns of the NASDAQ market index. In the last section, we conclude and discuss possible extensions. Proofs of the

theorems stated in the paper are gathered in the Appendix.

## 2 Regime Switching GARCH Model

The standard GARCH(1,1) model of Bollerslev (1986) is defined as the process

$$y_t = \mu_1 + \epsilon_t = \mu_1 + \sigma_t u_t \quad (1)$$

$$\sigma_t^2 = \omega_1 + \beta_1 \sigma_{t-1}^2 + \alpha_1 \epsilon_{t-1}^2, \quad (2)$$

where the error term  $u_t$  is i.i.d. with zero mean and unit variance. We assume for simplicity that the conditional mean is constant. The sum  $\beta_1 + \alpha_1$  measures the persistence of the volatility process. When this model is estimated using daily or higher frequency data, the estimate of this sum tends to be close to one, indicating that the volatility is highly persistent and the process may not be covariance-stationary. However it was argued that the high persistence may artificially result from regime shifts in the GARCH parameters over time, see Diebold (1986), Lamoureux and Lastrapes (1990), and Mikosch and Starica (2004), among others.

This motivates specifying a model that allows for regime-switching in the parameters. We define for each  $t$  an unobserved state variable  $s_t \in \{1, 2, \dots, n\}$ , which selects the model parameters with probability  $p_{jt} = \Pr(s_t = j | \mathfrak{S}_{t-1})$  where  $\mathfrak{S}_t$  is an information set available at time  $t$ , which includes  $(y_t, \sigma_t, \dots, y_1, \sigma_1)$ . Thus we define the RS-GARCH model as

$$y_t = \mu_{s_t} + \epsilon_t = \mu_{s_t} + \sigma_t u_t, \quad u_t \sim i.i.d.(0, 1) \quad (3)$$

$$\sigma_t^2 = \omega_{s_t} + \beta_{s_t} \sigma_{t-1}^2 + \alpha_{s_t} \epsilon_{t-1}^2 \quad (4)$$

$$p_{jt} = \Pr(s_t = j | \mathfrak{S}_{t-1}) = p_{jt}(y_{t-1}^2), \quad j = 1, 2, \dots, n, \quad (5)$$

where the function  $p_{jt}(\cdot)$  can be a logistic or exponential link function. This function depends on parameters not introduced explicitly at this stage and must be defined so that the probabilities are positive and sum to unity. In this way, the state probabilities are allowed to be time varying and the dynamics of the regimes can be determined endogenously. Obviously, the argument of the function  $p_{jt}(\cdot)$  can be different from  $y_{t-1}^2$ , e.g. the absolute value or higher lags, but it should be in the information set. This model was proposed by Gray (1996) who noted that it is analytically intractable, so that conditions for stability and existence

of moments had not yet been developed. The RS-GARCH model can be interpreted as a Markov chain with transition kernel that is a mixture of distributions. We state below simple assumptions for which this chain is geometrically ergodic and has finite moments. These results are based on Markov chain theory and use the concept of a drift function, see e.g. Meyn and Tweedie (1993) and Chan (1993).

In the RS-GARCH model, we assume that in the first regime ( $s_t = 1$ ) the process behaves locally as a strictly stationary process and its second moment exists. In other regimes ( $s_t > 1$ ), the conditional volatility can be integrated or even explosive, i.e. we impose no stability restrictions on the parameters of these regimes. Our assumption are the following:

**A1**  $u_t$  is i.i.d. and has a continuous positive density on  $\mathfrak{R}$ , with  $E(u_t) = 0$ ,  $\text{Var}(u_t) = 1$ .

**A2**  $\beta_j > 0$  and  $\alpha_j > 0$  for  $j = 1, 2, \dots, n$ .

**A3**  $p_{1t}(y_{t-1}^2) > 0$  and  $p_{1t}(y_{t-1}^2) \rightarrow 1$  as  $y_{t-1}^2 \rightarrow \infty$  for all  $t$ .

**A4**  $\beta_1 + \alpha_1 < 1$ , i.e. the first regime is stable.

Assumption A1 is standard and is satisfied for commonly used distributions for GARCH models, such as the normal and the Student with more than two degrees of freedom. Note that we could replace the requirement of continuity by lower semi-continuity. Assumption A2 is slightly stronger than the usual non-negativity conditions ( $\beta_j \geq 0$  and  $\alpha_j \geq 0$ ). It is needed in establishing the irreducibility of the process. This is not really restrictive since if  $\beta_j = \alpha_j = 0$  for all  $j$ , the model amounts to a static mixture model. In assumption A3, we postulate that the probability that  $s_t = 1$  is strictly positive. This implies that regardless of the state the process is in at time  $t - 1$ , there is always a positive probability that it will reach the stable state at time  $t$ . The assumption does not bound  $p_{1t}$  from above, so the process can spend its entire time in the stable regime. Furthermore, it is assumed that the stable process dominates the global process in the sense that the process collapses back to the stable regime when a big shock has occurred. These collapses relieve the pressure in the process, inducing stability as shown in the next theorem.

**Theorem 1** *Under assumptions A1-A4,  $y_t$  is geometrically ergodic. If the process is initiated from its invariant distribution, it is strictly stationary and  $\beta$ -mixing with exponential decay.*

The result means that there exists a unique invariant probability measure that solves the Chapman–Kolmogorov equation, and we can construct a stationary probability measure for

the process using the standard bottom-up method to obtain marginals, followed by application of the Kolmogorov extension theorem. Furthermore, the geometric ergodicity property implies that given any initial distribution, the process converges in total variation to its stationary measure with a uniform geometric rate. This implies that the process is exponentially  $\beta$ -mixing, so the autocovariance function tends to zero at an exponential rate (e.g., Davidson (1994), corollary 14.3). Unfortunately it is not possible to describe the stationary distribution in general. There exists different numerical techniques for solving this problem; see Tong (1996) for an overview. It is important to note that the process is strictly stationary regardless of the parameter values in the non-stable regimes, which means that the model can have periods of non-stationarity, but is globally stable.

The result on the existence of moments is given in the next theorem. We observe that the persistence parameters ( $\beta_1$  and  $\alpha_1$ ) in the stable GARCH regime are important in establishing the existence of moments. In particular, for  $k = 1$ , the conditions imposed in Theorem 1 are sufficient for covariance-stationarity of the process.

**Theorem 2** *Under assumptions A1-A3, if  $E(u_t^{2k}) < \infty$  and  $E(\beta_1 + \alpha_1 u_t^2)^k < 1$  for some  $k \geq 1$ ,  $y_t$  is geometrically ergodic and  $E(y_t^{2k}) < \infty$ .*

In the special case where the probabilities are constant, i.e.  $p_{jt} = p_j$  for all  $j$  and  $t$ , sufficient conditions for geometric ergodicity and existence of moment are given in Corollaries 1 and 3.

**Corollary 1** *Under assumptions A1-A2 and constant probabilities, if  $\sum_{j=1}^n p_j(\beta_j + \alpha_j) < 1$ , the results of Theorem 1 follow.*

This corollary implies that it is not necessary that the covariance-stationarity requirement,  $\beta_j + \alpha_j < 1$ , be satisfied for all the GARCH regimes, but it must be satisfied on average with respect to the probability distribution of the regimes. Note that high values of  $\beta_j + \alpha_j$  for the non-stable GARCH processes must match low enough probabilities for these regimes. This result is similar to the condition for strict stationarity of the mixed-normal GARCH model defined by Haas, Mittnik, and Paoletta (2004a); see also Francq, Roussignol, and Zakoian (2001).

**Corollary 2** *Under assumptions A1, A2, A4, and constant probabilities, if the probability of the stable regime is positive ( $p_1 > 0$ ) and the GARCH processes (4) for  $s_t > 1$  are integrated (IGARCH), the results of Theorem 1 follow.*

With more than two regimes, one may conceive that at least two different IGARCH processes co-exist with different persistence parameters that sum to one, but this does not seem very likely in practice. So in the perspective of applications, Corollary 2 may be relevant for a model with two regimes, one of them being an IGARCH. From this corollary, we see that in the special case where the explosive volatility processes are integrated, the RS-GARCH process is covariance-stationary regardless of the features of the state probabilities if there exists a real number  $\delta$  such that  $0 < \delta \leq p_1$ . When the state probabilities are time varying, they can be relatively high in tranquil periods regardless of the parameter values of the GARCH model in each regime, which allows one to deduce more information about the current state of the process given the available information. Furthermore, in the case of time varying probabilities, it is sufficient that the sum  $\beta_j + \alpha_j$  in one of the regimes is less than one, to ensure strict stationarity, which is not the case when the probabilities are fixed.

**Corollary 3** *Under assumptions A1-A2 and constant probabilities, if the probability of the stable regime is positive,  $E(u_t^{2k}) < \infty$  and  $\sum_{j=1}^n p_j E(\beta_j + \alpha_j u_t^2)^k < 1$  for some  $k \geq 1$ , the results of Theorem 2 follow.*

This corollary shows that, in terms of existence of moments, the condition  $E(\beta_j + \alpha_j u_t^2)^k < 1$  must hold only on average when probabilities are constant, whereas it must hold only for the stable regime when they are time varying.

### 3 Estimation

Given the current computing capability, the estimation of switching GARCH models by the maximum likelihood method is impossible, since the conditional variance depends on the whole past history of the state variables. Indeed, the conditional variance  $\sigma_t^2$  depends directly on the state variable  $s_t$  and  $\sigma_{t-1}^2$ , which itself depends on  $s_{t-1}$  and  $\sigma_{t-2}^2$ , and so on. To compute the likelihood function for the  $t$ -th observation we have to sum over  $n^t$  possible paths, which is computationally not feasible even for a relatively small sample size. In Figure 1, we illustrate the path dependence in a model with two regimes and zero means. The subscripts show the paths of the regimes: for example,  $\sigma_{2|1,2}^2$ , stands for the conditional variance at time 2, given that the process was in regime 1 and 2, respectively, at times 1 and 2. So, each conditional variance depends not just on the current regime, but on the entire history of the process. Cai (1994) and Hamilton and Susmel (1994) were the first to point out this problem. That is why

they apply the regime-switching approach only to a low order ARCH process. We tackle the estimation problem by Bayesian inference, which allows us to treat the latent state variables as parameters of the model and to construct the likelihood function assuming we know the states. This technique is called data augmentation, see Tanner and Wong (1987) for the basic principle and more details. In Section 3.1, we present the Bayesian algorithm for the case of two regimes, and in Section 3.2, we illustrate that it recovers correctly the parameters of a simulated data generating process.

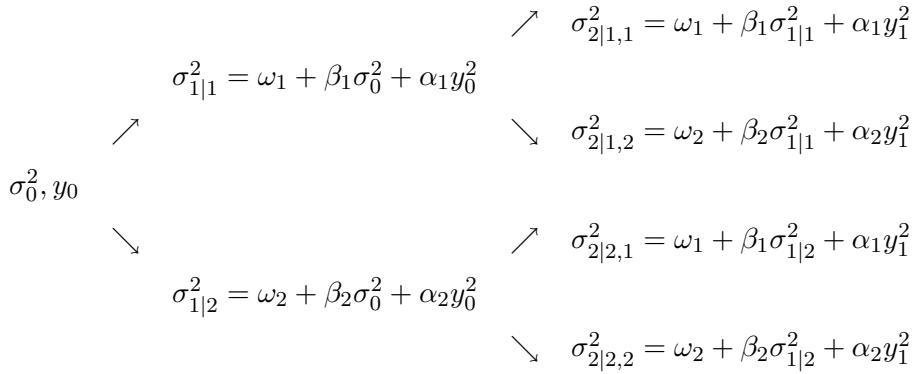


Figure 1: Path dependence in a RS-GARCH model with two states

### 3.1 Bayesian Inference

We explain the Bayesian algorithm for a RS-GARCH model with two regimes and normality of the error term  $u_t$ . The normality assumption is a natural starting point. A more flexible distribution, such as the Student distribution, could be considered, although one may be skeptical that this is needed since Gray (1996) reports large and imprecise estimates of the degrees of freedom parameters. The excess kurtosis property characterizing the standard GARCH model, is preserved in the two regime RS-GARCH model. Moreover if the two means differ ( $\mu_1 \neq \mu_2$ ), this model allows for time varying skewness (see e.g. McLachlan and Peel (2000)), a feature that can be implemented by the use of a skew distribution for  $u_t$  in the case of standard GARCH models, see e.g. Hansen (1994), Brooks, Burke, and Persaud (2002) and Wang, Fawson, Barrett, and McDonald (2001).

For the case of two regimes, the model is given by equations (3) and (4),  $s_t = 1$  indicating the stable regime and  $s_t = 2$  for the other regime, and a functional specification of either  $p_{1t}$

or  $p_{2t}$  (since  $p_{2t} = 1 - p_{1t}$ ). We specify

$$p_{1t} = [1 + \exp(\delta_0 + \delta_1 y_{t-1}^2)]^{-1}, \quad (6)$$

with  $\delta_1 < 0$ . Thus,  $p_{1t}$  tends to 1 as  $y_{t-1}^2$  tends to infinity, as required by assumption A3.

The model parameters consist of  $\delta = (\delta_0, \delta_1)'$ ,  $\mu = (\mu_1, \mu_2)'$ , and  $\theta = (\theta'_1, \theta'_2)'$ , where  $\theta_k = (\omega_k, \alpha_k, \beta_k)'$  for  $k = 1, 2$ . We denote by  $Y_t$  the vector  $(y_1 \ y_2 \ \dots \ y_t)$  and likewise  $S_t = (s_1 \ s_2 \ \dots \ s_t)$ . The joint density of  $y_t$  and  $s_t$  given the past information and the parameters is factorized as

$$f(y_t, s_t | \mu, \theta, \delta, Y_{t-1}, S_{t-1}) = f(y_t | s_t, \mu, \theta, Y_{t-1}, S_{t-1}) f(s_t | \delta, Y_{t-1}) \quad (7)$$

with

$$f(y_t | s_t, \mu, \theta, Y_{t-1}, S_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right) \quad (8)$$

(a normal density, where  $\sigma_t^2$  is a function of  $\theta$  through  $s_t$ , defined by equation (4)), and

$$f(s_t | \delta, Y_{t-1}) = p_{1t}^{2-s_t} (1 - p_{1t})^{s_t-1}, \quad (9)$$

(a Bernoulli distribution). The joint density of  $y = (y_1, y_2, \dots, y_T)$  and  $S = (s_1, s_2, \dots, s_T)$  given the parameters is then obtained by taking the product of the densities in (8) and (9) over all observations :

$$f(y, S | \mu, \theta, \delta) \propto \prod_{t=1}^T \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right) p_{1t}^{2-s_t} (1 - p_{1t})^{s_t-1}, \quad (10)$$

Since integrating this function with respect to  $S$  by summing over all paths of the state variables is numerically too demanding, we implement a Gibbs sampling algorithm that allows us to sample from the full conditional posterior densities of blocks of parameters given by  $\theta$ ,  $\mu$ ,  $\delta$ , and the elements of  $S$ . We define the prior densities of  $\theta$ ,  $\mu$ , and  $\delta$  when we explain the corresponding full conditional posterior densities in Sections 3.1.1-3.1.4.

### 3.1.1 Sampling $s_t$

The elements of  $S$  are not conditionally independent, because observations after time  $t$  are informative about  $s_t$  (this is related to the path dependence problem). Therefore,

$$\varphi(s_t | S_{\neq t}, \mu, \theta, \delta, y) \propto p_{1t}^{2-s_t} (1 - p_{1t})^{s_t-1} \prod_{j=t}^T \sigma_j^{-1} \exp\left(-\frac{(y_j - \mu_{s_j})^2}{2\sigma_j^2}\right) \quad (11)$$

where  $S_{\neq t}$  is the vector  $S$  without the element  $s_t$ , and  $\mu_{s_j}$  and  $\sigma_j^2$  depend on  $s_j$  since it defines the parameters operating at time  $j$ . Actually, to sample  $s_t$  we must condition only on the future state variables  $(s_{t+1}, s_{t+2}, \dots, s_T)$ , not on the past ones. To sample  $s_t$ , we evaluate (11) for  $s_t = 1$  and 2, normalize and draw as from a Bernoulli distribution.

### 3.1.2 Sampling $\theta$

Given a prior density  $\pi(\theta)$ ,

$$\varphi(\theta|S, \mu, \delta, y) \propto \pi(\theta) \prod_{t=1}^T \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right), \quad (12)$$

which does not depend on  $\delta$ . We sample  $\theta$  with the griddy-Gibbs sampler. The algorithm works as follows at iteration  $r + 1$ , given draws at iteration  $r$  denoted by the superscript  $(r)$  attached to the parameters:

1. Using (12), compute  $\kappa(\omega_1|S^{(r)}, \beta_1^{(r)}, \alpha_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$ , the kernel of the conditional posterior density of  $\omega_1$  given the values of  $S$ ,  $\beta_1$ ,  $\alpha_1$ ,  $\theta_2$ , and  $\mu$  sampled at iteration  $r$ , over a grid  $(\omega_1^1, \omega_1^2, \dots, \omega_1^G)$ , to obtain the vector  $G_\kappa = (\kappa_1, \kappa_2, \dots, \kappa_G)$ .
2. By a deterministic integration rule using  $M$  points, compute  $G_f = (0, f_2, \dots, f_G)$  where

$$f_i = \int_{\omega_1^1}^{\omega_1^i} \kappa(\omega_1|S^{(r)}, \beta_1^{(r)}, \alpha_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y) d\omega_1, \quad i=2, \dots, G. \quad (13)$$

3. Generate  $u \sim U(0, f_G)$  and invert  $f(\omega_1|S^{(r)}, \beta_1^{(r)}, \alpha_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$  by numerical interpolation to get a draw  $\omega_1^{(r+1)} \sim \varphi(\omega_1|S^{(r)}, \beta_1^{(r)}, \alpha_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$ .
4. Repeat steps 1-3 for  $\varphi(\beta_1|S^{(r)}, \omega_1^{(r+1)}, \alpha_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$ ,  
 $\varphi(\alpha_1|S^{(r)}, \omega_1^{(r+1)}, \beta_1^{(r+1)}, \theta_2^{(r)}, \mu^{(r)}, y)$ ,  $\varphi(\omega_2|S^{(r)}, \beta_2^{(r)}, \alpha_2^{(r)}, \theta_1^{(r+1)}, \mu^{(r)}, y)$ , etc.

Note that intervals of values for the elements of  $\theta_1$  and  $\theta_2$  must be defined. The choice of these bounds (such as  $\omega_1^1$  and  $\omega_1^G$ ) needs to be fine tuned in order to cover the range of the parameter over which the posterior is relevant. Over these intervals, the prior can be chosen as we wish, for example as uniform densities.

### 3.1.3 Sampling $\mu$

Given a prior density  $\pi(\mu)$ ,

$$\varphi(\mu|S, \theta, \delta, y) \propto \pi(\mu) \prod_{t=1}^T \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_{st})^2}{2\sigma_t^2}\right) \quad (14)$$

which does not depend on  $\delta$ . It is not possible to factorize this function into the product of a function that depends on  $\mu_1$  but not on  $\mu_2$ , and another one that depends on  $\mu_2$  but not on  $\mu_1$ . Moreover, since  $\sigma_t$  depends on  $\mu_1$  or  $\mu_2$  (depending on  $t$ ), the analytical form of the likelihood as a function of  $\mu_1$  or  $\mu_2$  is not a known density (e.g. a normal). Hence we must sample  $\mu_1$  and  $\mu_2$  jointly and numerically. We use the griddy-Gibbs sampler (a Metropolis sampler could also be used) and we factorize the prior as a product of two uniform densities.

### 3.1.4 Sampling $\delta$

Given a prior density  $\pi(\delta)$ ,

$$\varphi(\delta|S, \mu, \theta, y) \propto \pi(\delta) \prod_{t=1}^T p_{1t}^{2-st} (1 - p_{1t})^{st-1}, \quad (15)$$

which does not depend on  $\mu$  and  $\theta$ . We sample  $\delta = (\delta_0, \delta_1)'$  with the griddy-Gibbs sampler, so that given prior intervals, the prior can be chosen as we wish, for example as uniform densities.

## 3.2 Simulation Example

We have simulated a data generating process (DGP) corresponding to the model defined by equations (3)-(4)-(6), and  $u_t \sim N(0, 1)$ . The parameter values of the GARCH equations are actually equal to the posterior means obtained for the sample of the Nasdaq returns from 8/2/1971 to 22/12/1986, a sample of 4000 observations that precedes the sample used for the application of Section 4. One component is stable ( $\beta_1 + \alpha_1 = 0.892$ ) and the other is explosive ( $\beta_2 + \alpha_2 = 2.13$ ). The probability of the stable regime is given by  $1/[1 + \exp(-2.4 - 0.2y_{t-1}^2)]$  and fluctuates around 0.92. All the assumptions for stationarity and existence of moments of high order are satisfied. In Table 1, we report the summary statistics for 50,000 observations from this DGP, and in Figure 2, we show the 4,000 initial observations of the series, and based on the 50,000 observations, the estimated density of the data and the autocorrelations

of the squared data. The mean of the data is equal to 0.073, which is slightly smaller than  $0.92\mu_1 + (1 - 0.92)\mu_2$ , with  $\mu_1 = -\mu_2 = 0.09$ . The density is slightly skewed to the left, and its excess kurtosis, estimated to be 4.46, is much larger than 0.004, the excess kurtosis of the process when it is always in the stable regime. The ACF of the squared data is strikingly different from the ACF of the GARCH(1,1) process of the first regime, which would start at  $\rho_1 = 0.0126$  and decline according to  $\rho_n = 0.892\rho_{n-1}$ , so that  $\rho_{50} \simeq 0$ . The corresponding values in panel (c) of Figure 2 are 0.31 for  $\rho_1$  and very close to 0 for  $\rho_{50}$ . Thus, it is clear that, due to the explosive component, the ACF declines relatively slowly, although the decline rate is exponential. Said differently, a GARCH(1,1) process would have to be close to integrated to produce the excess kurtosis and the ACF shown in Figure 2.

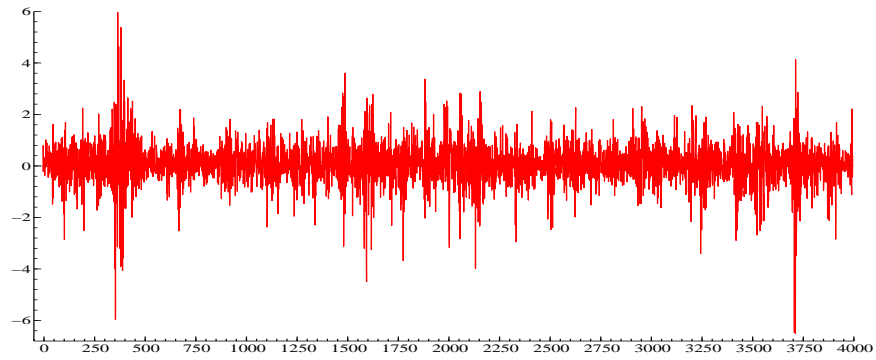
Table 1: Descriptive statistics for simulated data

Mean	0.073	Maximum	7.71
Standard deviation	0.805	Minimum	-6.64
Skewness	-0.042	Kurtosis	7.46

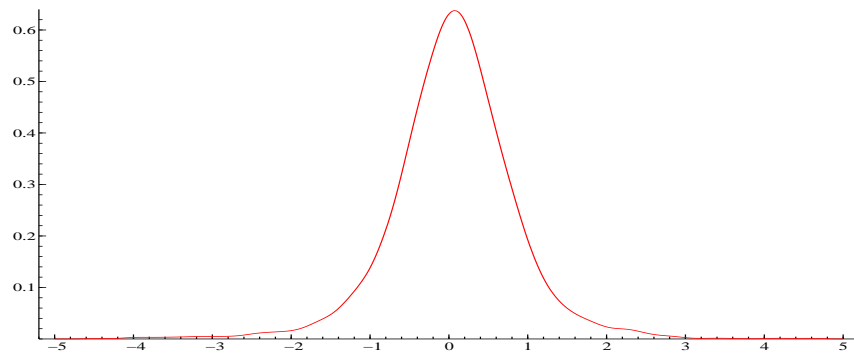
Statistics for 50,000 observations of the DGP defined in Table 2.

In Table 2, we report the posterior means and standard deviations for the model corresponding to the DGP, using the first 4,000 observations of the simulated data described above and shown in panel (a) of Figure 2. The results are in the column with the header RS-GARCH. The bounds for the prior density of each parameter are reported in the second column of Table 2 with the DGP values. Thus, these bounds were used for the integrations in the gridy-Gibbs sampler. The number of iterations of the Gibbs sampler was set to 10,000, but the initial 4,000 draws were discarded, because after these the sampler seems to have converged (based on cumsum diagrams not reproduced to save space). Thus the posterior moments are based on 6,000 dependent draws of the posterior distribution. The posterior means are close to the DGP values for the sample we used (4,000 observations). We also find that the posterior standard deviations are much larger for the parameters related to the second regime than to the first one, which is consistent with the fact that about 8 per cent of the observations can be considered as "belonging" to the second regime. The computing time for generating 10,000 draws was about 100 hours using a 3.2 Ghz Intel Xeon processor.

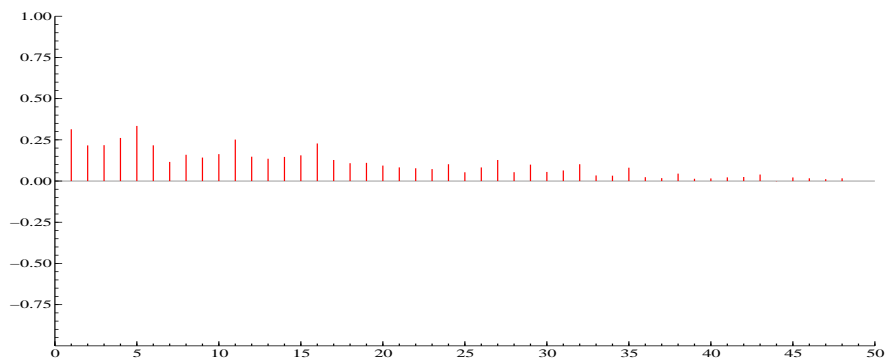
The last column of Table 2 contains the results for a GARCH(1,1) model with  $N(0, 1)$



(a) Sample path (4,000 observations)



(b) Kernel density (50,000 observations)



(c) Correlogram of squared data (50,000 observations)

Figure 2: Graphs for simulated data for DGP defined in Table 2

Table 2: Posterior means and standard deviations (simulated DGP)

	DGP values (prior bounds)	RS-GARCH	N-GARCH
$\omega_1$	0.012 (0.001,0.024)	0.013 (0.0052)	0.023 (0.0027)
$\beta_1$	0.88 (0.780,0.99)	0.88 (0.039)	0.80 (0.011)
$\alpha_1$	0.012 (0.001,0.022)	0.013 (0.0071)	0.17 (0.011)
$\omega_2$	0.12 (0.005,0.37)	0.13 (0.094)	-
$\beta_2$	1.90 (1.05,3)	2.19 (0.39)	-
$\alpha_2$	0.13 (0.01,0.5)	0.23 (0.15)	-
$\mu_1$	0.09 (0.045 0.11)	0.076 (0.0074)	0.074 (0.0077)
$\mu_2$	-0.09 (-0.3 0.25)	0.039 (0.069)	-
$\delta_0$	-2.40 (-3.8,-1.4)	-2.67 (0.42)	-
$\delta_1$	-0.200 (-0.7 0.1)	-0.30 (0.20)	-

Posterior standard deviations under posterior means.  
Sample of 4,000 observations from DGP.  
DGP: equations (3)-(4)-(6) with  $N(0,1)$  distribution.  
RS-GARCH: same model as DGP.  
N-GARCH: equations (1)-(2), and  $N(0,1)$  distribution.

errors, for the same sample of 4,000 observations. These results are based on 6,000 draws obtained by a griddy-Gibbs sampler after discarding 4,000 draws as a warm-up sample. These results illustrate that the posterior mean of the persistence effect  $(\alpha_1 + \beta_1)$ , equal to 0.97, is pushed towards 1 due to the wrong specification of the model. Forgetting the second component forces the estimates to adjust to fit the persistence in the squared data illustrated in panel (c) of Figure 2.

## 4 Application

We use the NASDAQ daily percentage returns from 23/12/1986 to 28/10/2002 (4,000 observations) for estimation. Figure 3 displays the sample path, the kernel density, and the correlogram of the squared returns. We observe a strong persistence in the squared returns, negative skewness, and a high kurtosis, see also Table 3.

Table 3: Descriptive statistics for NASDAQ daily returns

Mean	0.033	Minimum	-12.05
Standard deviation	1.52	Maximum	13.26
Skewness	-0.24	Kurtosis	10.93
Sample period: 23/12/1986 to 28/10/2002 (4,000 observations)			

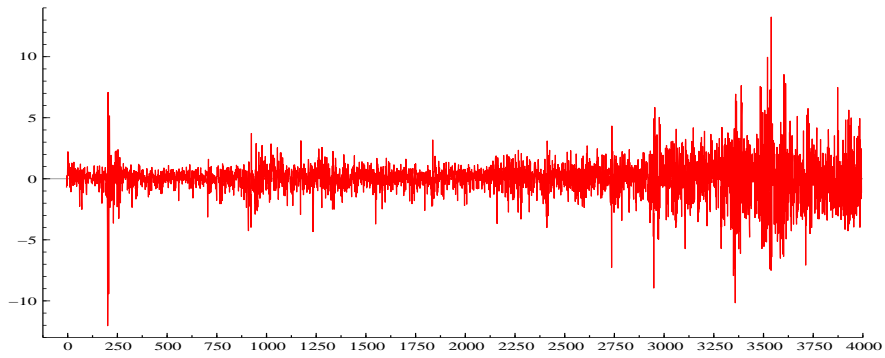
In Table 4, we report the posterior means and standard deviations from the estimation of different models using the estimation sample. The estimated models include the two-regime RS-GARCH model defined by equations (3), (4), and (6), the standard GARCH model defined by equations (1) and (2) with  $u_t$  distributed as  $N(0, 1)$ , the same model with  $u_t \sim t_\nu$  (Student distribution with  $\nu$  degrees of freedom), and the MN-GARCH(2,2) model of Haas, Mittnik, and Paoletta (2004a), defined by

$$y_t = \begin{cases} \mu_1 + \sigma_{1t}u_t, & \text{with probability } p \\ \mu_2 + \sigma_{2t}u_t, & \text{with probability } 1 - p, \end{cases} \quad (16)$$

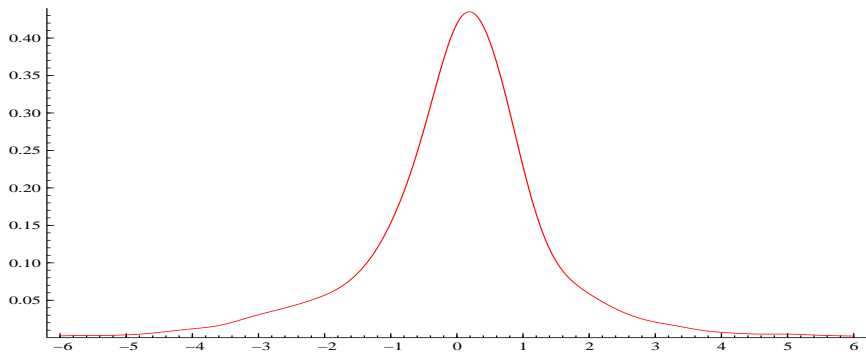
$$\sigma_{1t}^2 = \omega_1 + \beta_1\sigma_{1,t-1}^2 + \alpha_1y_{t-1}^2, \quad (17)$$

$$\sigma_{2t}^2 = \omega_2 + \beta_2\sigma_{2,t-1}^2 + \alpha_2y_{t-1}^2, \quad (18)$$

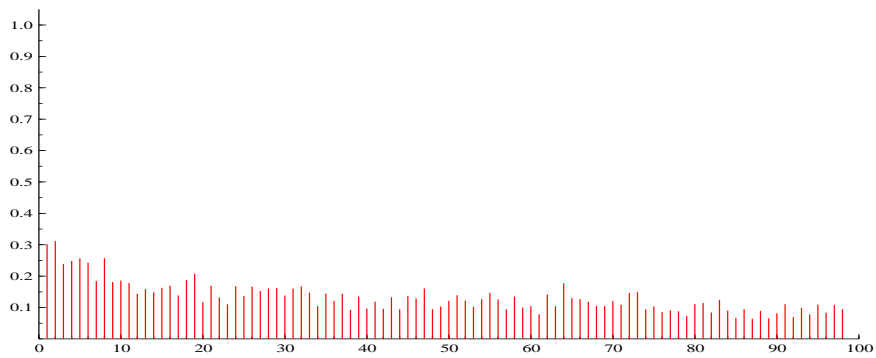
$$u_t \sim i.i.d. N(0, 1). \quad (19)$$



(a) Sample path (4,000 observations)



(b) Kernel density



(c) Correlogram of squared data

Figure 3: NASDAQ daily returns from 23/12/1986 to 28/10/2002

Equivalently, the distribution of  $y_t$  is the Gaussian mixture  $pN(\mu_1, \sigma_{1t}^2) + (1 - p)N(\mu_2, \sigma_{2t}^2)$ . For identification we impose that  $p$  is larger than 0.5. Bauwens and Rombouts (2005) show how to estimate this model by Bayesian inference. However, we do not impose that the mixture mean is equal to zero, hence  $\mu_1$  and  $\mu_2$  are not linked by imposing that  $p\mu_1 + (1 - p)\mu_2$  is equal to 0. Note that for the MN-GARCH model with a fixed probability, we report the results for  $p = (1 + \exp(\delta_0))^{-1}$  rather than  $\delta_0$ . A few comments about the estimation results follow.

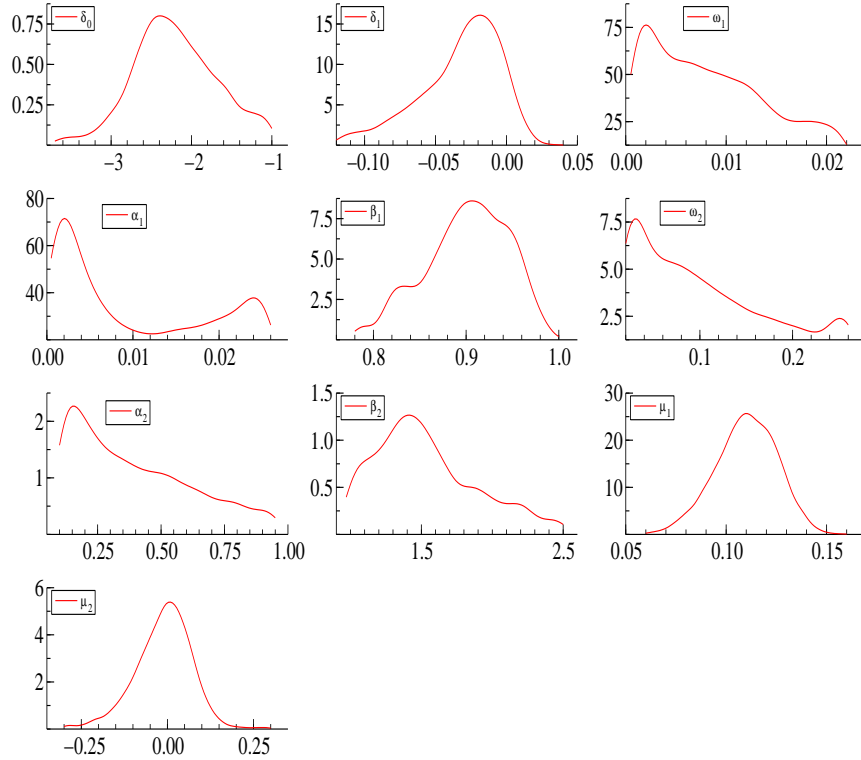


Figure 4: Posterior densities for the RS-GARCH model (NASDAQ daily returns)

1) RS-GARCH: The marginal posterior densities are shown in Figure 4. The intervals over which the densities are drawn are the prior intervals. The intervals for the GARCH parameters were chosen to avoid negative values, and by trial and error so as to avoid too much truncation. The density of  $\alpha_1$  is clearly bimodal, like in the simulation example of the previous section. The first regime is stable, with the posterior mean of  $\beta_1 + \alpha_1$  equal to 0.92 and

Table 4: Posterior means and standard deviations (NASDAQ daily returns)

	RS-GARCH	MN-GARCH	$N$ -GARCH	$t$ -GARCH
$\omega_1$	0.0073 (0.0052)	0.0008 (0.0007)	0.0217 (0.0033)	0.0073 (0.0028)
$\beta_1$	0.907 (0.038)	0.922 (0.010)	0.848 (0.012)	0.887 (0.020)
$\alpha_1$	0.0119 (0.0086)	0.057 (0.0087)	0.148 (0.013)	0.082 (0.016)
$\omega_2$	0.108 (0.070)	0.121 (0.049)	-	-
$\beta_2$	1.451 (0.32)	0.779 (0.157)	-	-
$\alpha_2$	0.428 (0.237)	0.631 (0.055)	-	-
$\mu_1$	0.108 (0.014)	0.165 (0.018)	0.082 (0.0098)	0.105 (0.013)
$\mu_2$	-0.005 (0.079)	-0.641 (0.118)	-	-
$\delta_0$ or $p$	$\delta_0$ : -2.218 (0.448)	$p$ : 0.845 (0.029)	-	-
$\delta_1$	-0.032 (0.0273)	-	-	-
$\nu$	-	-	-	7.09 (0.91)

Posterior standard deviation in parenthesis under the corresponding posterior mean.

Sample period: 23/12/1986 to 28/10/2002 (4,000 observations).

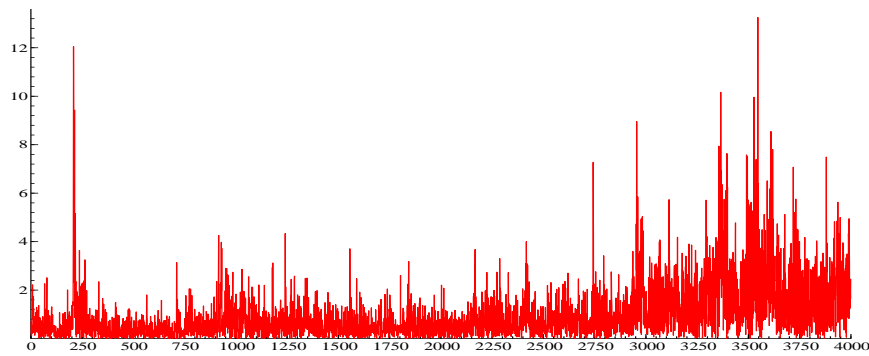
RS-GARCH: equations (3)-(4)-(6) and  $N(0, 1)$  distribution.

MN-GARCH: equations (16)-(19).

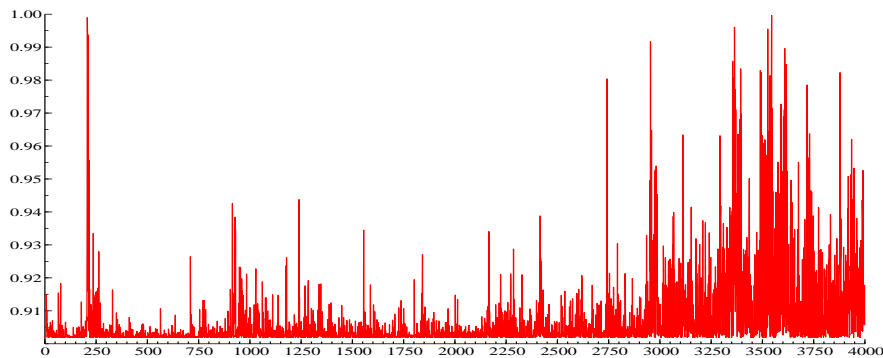
$N$ -GARCH: equations (1)-(2) and  $N(0, 1)$  distribution.

$t$ -GARCH: equations (1)-(2) and  $t_\nu$  distribution.

the second one is explosive, with the posterior mean of  $\beta_2 + \alpha_2$  equal to 1.88. The statistical evidence for a time-varying probability is not strong, since the posterior mean of  $\delta_1$  is equal to  $-0.032$  whereas the standard deviation is equal to 0.027. Computing the time-varying probability of the first regime at the posterior mean of  $\delta_0$  and  $\delta_1$ , we get the value 0.90 if  $y_{t-1} = 0$  and 0.999 if  $|y_{t-1}| = 13.26$  (the largest observed absolute return, observed on January 3, 2001). Figure 5 displays the path of these probabilities in parallel with the absolute returns. In fact, the probability of the stable regime increases when the return becomes more volatile the day before.



(a) Absolute returns



(b) Probabilities of stable regime

Figure 5: NASDAQ daily returns from 23/12/1986 to 28/10/2002

2) MN-GARCH: The posterior mean of the mixture expectation  $p\mu_1 + (1 - p)\mu_2$  is equal to 0.042, which is not far from the data mean. The posterior mean of the probability parameter  $p$  is equal to 0.85, which is smaller than for the RS-GARCH model. The GARCH parameters

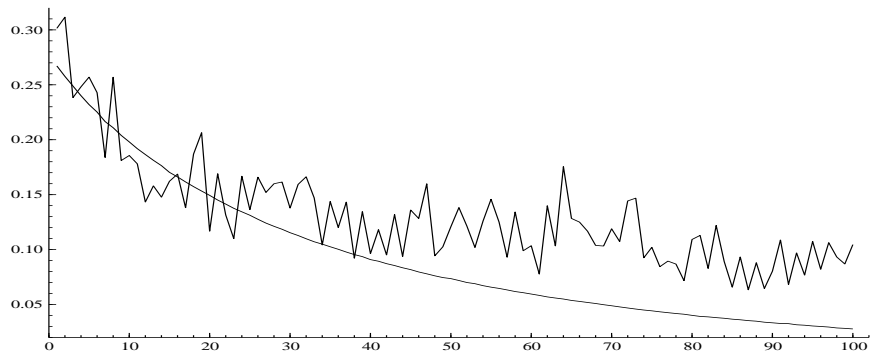
of both models are either very different  $(\omega_1, \alpha_1, \alpha_2, \beta_2)$  or very similar  $(\beta_1, \omega_2)$ , and the sum  $\alpha_2 + \beta_2$  is estimated to be equal to 1.88 in the RS-GARCH model and 1.41 in the MN-GARCH one. The weak stationarity condition for this model, see Haas et al. (2004), is that  $[(1 - \beta_1 - \alpha_1)p/(1 - \beta_1) + (1 - \beta_2 - \alpha_2)(1 - p)/(1 - \beta_2)](1 - \beta_1)(1 - \beta_2)$  is positive. The mean of this function of the parameters computed for all Gibbs draws is equal to 0.0017 with a standard deviation of 0.0011. The posterior probability that the unconditional variance does not exist amounts to 0.05 (i.e. the function is negative for five per cent of the Gibbs draws).

3) Normal-GARCH: The posterior mean of the persistence parameter  $\beta_1 + \alpha_1$  is equal to 0.996, which makes it almost an IGARCH model.

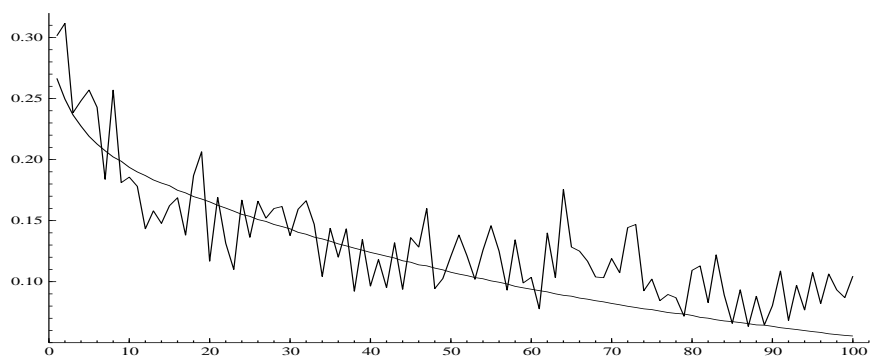
4)  $t$ -GARCH: The posterior mean of the degrees of freedom parameter is close to 7. The relevant persistence parameter for this model is  $\beta_1 + \alpha_1\nu/(\nu - 2)$ . Evaluated at the posterior means, this is equal to 1.002 indicating again that this simple model gives the illusion that the volatility is integrated.

To compare informally the models, we evaluate their ability to fit the sample variance, autocorrelations, skewness and kurtosis coefficients by the unconditional moments they imply. For a given model, each unconditional moment is a function of the model parameters. For the RS-GARCH model, since these functions are not known analytically, we compute them for each Gibbs draw of the posterior, using a simulated sample of 10,000 observations of the model. These simulations generate a sample of the posterior distribution of the function of interest (an implied moment), and we use the mean of the sample as an estimate of the implied moment (actually, this is the posterior mean of the implied moment). In doing this, we discarded a few Gibbs draws that generated extremely high values of the implied variance (specifically those for which it is larger than 8). The results are displayed in Table 5.

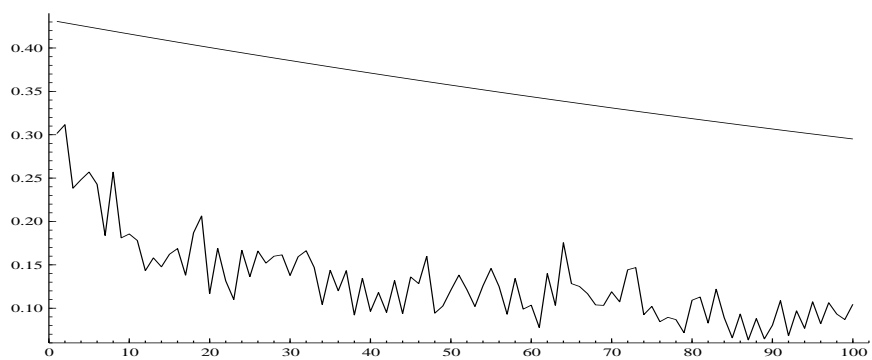
The RS-GARCH model fits the variance and the kurtosis of the data reasonably well but the model has some difficulties to generate the skewness available in the NASDAQ data. The MN-GARCH model performs better with respect to the skewness but it implies a much too high kurtosis. Similar results for the normal-GARCH and  $t$ -GARCH models are not reported, since the implied moments do not exist with high posterior probability.



(a) RS-GARCH



(b) MN-GARCH



(c) Normal-GARCH

Figure 6: Implied (smooth line) and empirical autocorrelation functions

Table 5: Implied and empirical moments (NASDAQ daily returns)

	Empirical	RS-GARCH	MN-GARCH
Mean	0.033	0.097	0.042
Variance	2.31	2.38	2.26
Skewness	-0.24	-0.016	-0.34
Kurtosis	10.93	12.77	20.86

Figure 6 gives an idea of the fit of the autocorrelation structure. The RS- and MN-GARCH models perform well until the order 40. After this order until about order 90 the MN-GARCH model remains closer to the sample autocorrelations. Finally, after lag 90, the implied autocorrelations of the MN-GARCH model also fade out to zero while the sample autocorrelations remain close to 0.1. This is due to the fact that the autocorrelations of the NASDAQ series do not seem to decrease at an exponential rate whereas the models can only imply either an exponential decrease rate or non-existing second order moments. We also provide in the last panel of Figure 6 the implied autocorrelation structure of the normal-GARCH model. Ding and Granger (1996) show that even if the fourth moment does not exist, the  $k$ th order autocorrelation can still be approximated by  $(\alpha_1 + \frac{1}{3}\beta_1)(\alpha_1 + \beta_1)^k$ . As expected, the normal-GARCH model performs very badly. The same conclusion holds for the  $t$ -GARCH model.

## 5 Conclusion

We have developed a regime-switching univariate GARCH model with a time-varying probability of switching between a stable volatility regime and an explosive one. We provide sufficient conditions for the stationarity of the process and the existence of moments. These results are original even if the probabilities are constant. The model cannot be estimated by the ML method because of the path dependence problem that has previously prevented estimation of such models, unless one was willing to use Gray's or Dueker's approximations without knowing the quality of the approximations. Thus, further research could be devoted to comparing the results of the Bayesian estimation algorithm we have developed with the approximate ML solutions. Another idea to improve upon these approximate ML estimators

is to use indirect inference since the RS-GARCH model can be easily simulated.

Future research about the regime-switching GARCH model could be focussed on estimating the model with other data series, and on in-sample and out-sample comparisons of the RS-GARCH models with other GARCH models. The specification of the model could also be extended, for example the time-varying probability could be a function of the lagged conditional variance rather than the lagged return, although this would complicate the estimation algorithm and require additional work to prove stationarity of the process.

## Appendix

To prove Theorems 1 and 2, we write the model in its Markovian state space representation. We use the notation  $\sigma_t^2 = h_{t-1}$  to make it clear that  $\sigma_t^2$  is a function of information dated at  $t - 1$  or earlier, not of information dated at  $t$ . Thus we write

$$y_t = \mu_j + \sqrt{h_{t-1}}u_t, \text{ with probability } p_i(y_{t-1}^2) \quad (20)$$

$$h_t = \omega_j + \beta_j h_{t-1} + \alpha_j \epsilon_t^2 = \omega_j + (\beta_j + \alpha_j u_t^2)h_{t-1} \quad (21)$$

with probability  $p_j \left( (\mu_j + \sqrt{h_{t-1}}u_t)^2 \right)$ .

**Proof of Theorem 1:** The process  $\{y_t, h_t\}$  forms a homogeneous Markov chain with state space given by  $(S, \mathfrak{S}, \lambda)$ , where  $S = \{(y, h) \in \mathfrak{R} \times \mathfrak{R}_+ : h \geq \omega_1 + \beta_1 \bar{h} + \alpha_1 (y - \mu_1)^2\}$ , and  $\bar{h} = \omega_1 / (1 - \beta_1)$ ,  $\mathfrak{S}$  is the Borel  $\sigma$ -algebra on  $\mathfrak{R} \times \mathfrak{R}_+$  restricted to  $S$  ( $\mathfrak{R}_+ \equiv (0, \infty)$ ) and  $\lambda$  denotes the usual Lebesgue measure on  $(S, \mathfrak{S})$ . Let  $P^m((y_0, h_0), A) = P(\{(y_t, h_t) \in A | y_{t-m} = y_0, h_{t-m} = h_0\})$  denote the probability that  $\{y_t, h_t\}$  moves from  $(y_0, h_0)$  to the set  $A \in \mathfrak{S}$  in  $m$  steps. In order to establish geometric ergodicity of the Markov chain, we first show that the process is  $\lambda$ -irreducible and we take similar steps as Zhang, Russell, and Tsay (2001). For irreducibility it is sufficient to show that  $\sum_{k=1}^{\infty} P^k((y_0, h_0), A) > 0$  for all  $(y_0, h_0) \in S$  and all Borel measurable sets  $A \in \mathfrak{S}$  with positive Lebesgue measure (Chan (1993)). We start by showing that for a given  $(y_0, h_0) \in S$  we can reach any  $(y, h) \in A$  by a finite number of steps. Let  $\tilde{h} = (h - \omega_1 - \alpha_1 (y - \mu_1)^2) / \beta_1$  and  $\varsigma = \tilde{h} - \bar{h}$ . Since  $h > \bar{h}$ , we have  $\varsigma > 0$ . So, there exists a positive integer  $m = \min\{j \geq 1 : \bar{h} + \varsigma + (\tilde{h} - \bar{h} - \varsigma)\beta_1^{-j} > \omega_1 - \beta_1 h_0\}$ , such that under the first regime, we can reach the point  $(y, h)$  through the  $m + 1$  intermediate steps  $z = \{(y_i, h_i)\}_{i=1}^{m+1}$ , where  $h_i = \bar{h} + \varsigma / 2 + (\tilde{h} - \bar{h} - \varsigma / 2)\beta_1^{i-(m+1)}$ ,  $y_1 = \mu_1 + ((h_1 - \omega_1 - \beta_1 h_0) / \alpha_1)^{0.5}$  and  $y_i = \mu_1 + (\varsigma(1 - \beta_1) / 2\alpha_1)^{0.5}$  for  $i > 1$ . The continuity of the state space equation implies

that there exists an open set  $D$  around any  $(y, h) \in A$  such that for  $D^* = \{(z, y) : (y, h) \in D \cap A\}$ ,  $\lambda(D^*) > 0$  and given assumption A3 there exists a positive real number for which  $\delta \leq p_1(\cdot)$ , hence

$$\begin{aligned} \sum_{k=1}^{\infty} P^k((y_0, h_0), A) &\geq P^{m+2}((y_0, h_0), D \cap A) \geq \int_{D^*} \prod_{i=0}^{m+1} p_1(y_i^2) f(y_{i+1}/h_i) dz dy \geq \\ &\geq \delta^{m+2} \int_{D^*} \prod_{i=0}^{m+1} f(y_{i+1}/h_i) dz dy > 0. \end{aligned}$$

Therefore the chain is  $\lambda$ -irreducible. It is straightforward to prove that, for each bounded continuous function  $g$  on  $S$ ,  $E(g(y_t, h_t) | y_{t-1} = y, h_{t-1} = h)$  is continuous in  $(y, h)$ , so the process is a  $\lambda$ -irreducible Feller chain. Given this result it is sufficient to verify the drift condition stated in Feigin and Tweedie (1985). As drift function we use  $V(y, h) = 1 + \bar{\eta}y^2 + h$ ,  $\bar{\eta} = \eta - \beta_1 - \alpha_1$  where  $\eta$  is some positive number which satisfies  $\beta_1 + \alpha_1 < \eta < 1$  and the test set is given by  $C = \{(y, h) \in S : h + y^2 \leq c\}$ , where  $c$  is a positive number to be determined below. Thus,

$$\begin{aligned} E(V(y_t, h_t) | h_{t-1} = h, y_{t-1} = y) &= 1 + \bar{\eta}h + M_1 + (\beta_1 + \alpha_1)h + \\ &+ h \sum_{i=1}^n p_i(y^2) \sum_{j \neq 1} E \left[ p_j((\mu_i + \sqrt{h} u_t)^2) [(\beta_j + \alpha_j u_t^2) - (\beta_1 + \alpha_1 u_t^2)] \right] \\ &= 1 + h \cdot \left\{ \eta + \frac{M}{h} + \sum_{j \neq 1} E \left[ p_j((\mu_i + \sqrt{h} u_t)^2) \cdot [(\beta_j + \alpha_j u_t^2) - (\beta_1 + \alpha_1 u_t^2)] \right] \right\} \end{aligned}$$

where,  $M = \sum_{i,j} \left\{ \bar{\eta} \mu_i^2 p_i(y^2) + p_i(y^2) E \left[ \omega_j p_j((\mu_i + \sqrt{h} u_t)^2) \right] \right\}$ .

Assumption A3 implies that

$$\frac{M}{h} + \sum_{i=1}^n p_i(y^2) \sum_{j \neq 1} E \left[ p_j((\mu_i + \sqrt{h} u_t)^2) \cdot [(\beta_j + \alpha_j u_t^2) - (\beta_1 + \alpha_1 u_t^2)] \right] \rightarrow 0 \text{ as } h \rightarrow \infty.$$

Since  $E(V(y_t, h_t) | h_{t-1} = h, y_{t-1} = y)$  is continuous and hence bounded on a compact set and  $h < V(y, h)$ , we can choose  $c$  such that  $E(V(y_t, h_t) | h_{t-1} = h, y_{t-1} = y) \leq \eta \cdot v(y, h) + a \cdot 1_C(y, h)$  for some  $a < \infty$  and for all  $(y, h)$ , hence the drift criterion is satisfied. By theorem 2 of Feigin and Tweedie (1985)  $\{y_t, h_t\}$  is geometric ergodic and so is  $\{y_t\}$ . The chain is also V-geometric mixing (Meyn and Tweedie (1993), theorem 16.1.5) which is equivalent to the requirement that the process is  $\beta$ -mixing with exponential decay.

**Proof of Theorem 2:** In this case it is sufficient to consider the drift function  $V(y, h) = 1 + (\bar{\eta}/\Delta)y^{2k} + h^k$  where  $\Delta = E(u_t^{2k})$ ,  $\bar{\eta} = \eta - E(\beta_1 + \alpha_1 u_t^2)^k$  and  $\eta$  is some positive number which satisfies  $E(\beta_1 + \alpha_1 u_t^2)^k < \eta < 1$  and the test set is given by  $C = \{(h, y) \in S : h + y^2 \leq c\}$ . After some tedious calculations, we have

$$\begin{aligned} E(V(y_t, h_t)|h_{t-1} = h, y_{t-1} = y) &= 1 + \bar{\eta}h^k + O(h^{k-1}) + E(\beta_1 + \alpha_1 u_t^2)^k h^k \\ &+ \sum_{i=1}^n p_i(y^2) \left\{ O(h^{k-1}) + h^k \sum_{j \neq 1} E \left[ p_j((\mu_i + \sqrt{h}u_t^2) \cdot [(\beta_j + \alpha_j u_t^2)^k - (\beta_1 + \alpha_1 u_t^2)^k]) \right] \right\} \\ &= 1 + h^k \cdot \left\{ \eta + \frac{O(h^{k-1})}{h^k} + \sum_{j \neq 1} E \left[ p_j((\mu_i + \sqrt{h}u_t^2) \cdot [(\beta_j + \alpha_j u_t^2)^k - (\beta_1 + \alpha_1 u_t^2)^k]) \right] \right\} \end{aligned}$$

Assumption A3 and the boundedness of the moment of the error term imply that

$$\frac{O(h^{k-1})}{h^k} + \sum_{j \neq 1} E \left[ p_j((\mu_i + \sqrt{h}u_t^2) \cdot [(\beta_j + \alpha_j u_t^2)^k - (\beta_1 + \alpha_1 u_t^2)^k]) \right] \rightarrow 0 \text{ as } h \rightarrow \infty.$$

Using the same arguments as in Theorem 1, we can show that there exists a compact set  $C$  such that  $E(V(y_t, h_t)|h_{t-1} = h, y_{t-1} = y) \leq \eta \cdot V(y, h) + a \cdot 1_C(y, h)$  for some  $a < \infty$  and for all  $(y, h)$  which implies the desired result.

**Proof of Corollary 1:** We follow the same steps as in Theorem 1 and set  $\bar{\eta} = \eta - \sum_{j=1}^n p_j(\beta_j + \alpha_j)$  where  $\eta$  satisfies  $\sum_{j=1}^n p_j(\beta_j + \alpha_j) < \eta < 1$ .

**Proof of Corollary 2:** It is straightforward given Corollary 1.

**Proof of Corollary 3:** We follow the same steps as in Theorem 2 and set  $\bar{\eta} = \eta - \sum_{j=1}^n p_j E(\beta_j + \alpha_j u_t^2)^k$  where  $0 < \bar{\eta} < \eta < 1$ .

## References

- BAUWENS, L., C. BOS, AND H. VAN DIJK (1999): “Adaptive Polar Sampling with an Application to a Bayes Measure of Value-at-Risk,” Tinbergen Institute, Discussion Paper TI 99-082/4.
- BOLLEN, S., N. GRAY, AND R. WHALEY (2000): “Regime-Switching in Foreign Exchange Rates: Evidence From Currency Option Prices,” *Journal of Econometrics*, 94, 239–276.
- BOLLERSLEV, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.

- BOLLERSLEV, T., AND J. WOOLDRIDGE (1992): “Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances,” *Econometric Reviews*, 11, 143–172.
- BROOKS, C., S. BURKE, AND G. PERSAND (2002): “Autoregressive Conditional Kurtosis,” ISMA Centre Discussion Papers in Finance Series 2002-5.
- CAI, J. (1994): “Markov Model of Unconditional Variance in ARCH,” *Journal of Business and Economics Statistics*, 12, 309–316.
- CHAN, K. (1993): *A Review of Some Limit Theorems of Markov Chains and Their Applications*. In H.Tong, Editor, *Dimension Estimation and Models*, World Scientific Publishing, Singapore and River Edge, NJ, USA.
- DAVIDSON, J. (1994): “Stochastic Limit Theory,” New York, Oxford University Press.
- DIEBOLD, F. (1986): “Comment on Modeling the Persistence of Conditional Variances,” *Econometric Reviews*, 5, 51–56.
- DING, Z., AND C. GRANGER (1996): “Modeling Volatility Persistence of Speculative Returns: A New Approach,” *Journal of Econometrics*, 73, 185–215.
- DUEKER, M. (1997): “Markov Switching in GARCH Processes in Mean Reverting Stock Market Volatility,” *Journal of Business and Economics Statistics*, 15, 26–34.
- FEIGIN, P., AND R. TWEEDIE (1985): “Random Coefficient autoregressive processes: A Markov chain analysis of stationary and finiteness of moments,” *Journal of Time Series Analysis*, 6, 1–14.
- FRANCQ, C., M. ROUSSIGNOL, AND J. ZAKOIAN (2001): “Conditional heteroskedasticity driven by hidden Markov chains,” *Journal of Time Series Analysis*, 22, 197–220.
- GRAY, S. (1996): “Modeling the conditional distribution of interest rates as a regime-switching process,” *Journal of Financial Economics*, 42, 27–62.
- HAAS, M., S. MITTNIK, AND M. PAOLELLA (2004a): “Mixed Normal Conditional Heteroskedasticity,” *Journal of Financial Econometrics*, 2, 211–250.

- (2004b): “A New Approach to Markov-Switching GARCH Models,” *Journal of Financial Econometrics*, 2, 493–530.
- HAMILTON, J., AND R. SUSMEL (1994): “Autoregressive Conditional Heteroskedasticity and Changes in Regime,” *Journal of Econometrics*, 64, 307–333.
- HANSEN, B. (1994): “Autoregressive Conditional Density Estimation,” *International Economic Review*, 35, 705–730.
- KIM, C. (1994): “Dynamic Linear Models with Markov Switching,” *Journal of Econometrics*, 60, 1–22.
- KLAASSEN, F. (2002): “Improving GARCH Volatility Forecasts with Regime-Switching GARCH,” Universiteit Tubingen, Working Paper.
- LAMOUREUX, C., AND W. LASTRAPES (1990): “Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects,” *Journal of Finance*, 45, 221–229.
- LANNE, M., AND P. SAIKKONEN (2003): “Modeling the U.S. Short Term Interest Rate by Mixture of Autoregressive Processes,” *Journal of Financial Econometrics*, 1, 96–125.
- MCLACHLAN, G., AND D. PEEL (2000): *Finite Mixture Models*. Wiley Interscience, New York.
- MEYN, S., AND R. TWEEDIE (1993): “Markov Chains and Stochastic Stability,” London, Springer Verlag.
- 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.
- SCHWERT, G. (1989): “Why Does Stock Market Volatility Change Over Time?,” *Journal of Finance*, 44, 1115–1153.
- TANNER, M., AND W. WONG (1987): “The Calculation of the Posterior Distributions by Data Augmentation,” *Journal of the American Statistical Association*, 82, 528–540.
- TONG, H. (1996): “Non-linear Time series: A Dynamical System Approach,” Oxford, Oxford University Press.

- VLAAR, P., AND F. PALM (1993): “The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps,” *Journal of Business and Economic Statistics*, 11, 351–360.
- WANG, K., C. FAWSON, C. BARRETT, AND J. McDONALD (2001): “A Flexible Parametric GARCH Model with an Application to Exchange Rates,” *Journal of Applied Econometrics*, 16, 521–536.
- WONG, C., AND W. LI (2001): “On a Mixture Autoregressive Conditional Heteroscedastic Model,” *Journal of the American Statistical Association*, 96, 982–995.
- ZHANG, M. Y., J. RUSSELL, AND R. TSAY (2001): “A nonlinear autoregressive conditional duration model with applications to financial transaction data,” *Journal of Econometrics*, 104, 179–207.