

CORE DISCUSSION PAPER

2006/11

REGIME SWITCHING GARCH MODELS

Luc Bauwens¹, Arie Preminger,² and Jeroen Rombouts³

February 20, 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 stationarity and existence of moments. Because of path dependence, maximum likelihood estimation is infeasible. 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.

¹CORE and Department of Economics, Université catholique de Louvain.

²CORE, Université catholique de Louvain.

³HEC Montréal, CIRANO, CIRPEE and CREF.

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 the 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 Paoletta (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 Paoletta (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. 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 RS-GARCH 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 infeasible. 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 with 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 Haas, Mittnik, and Paoletta (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 and predictive densities. Therefore our solution is rigorous in the sense that we estimate 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 and used for prediction 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

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

$$\sigma_t^2 = \alpha_1 + \beta_1 \sigma_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2, \quad (2)$$

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 $\beta_1 + \gamma_1$ 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 process is highly persistent and 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, Diebold (1986), Lamoureux and Lastrapes (1990), and Mikosch and Starica (2004), among others.

This motivates specifying the RS-GARCH 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

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

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

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

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 first 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. Here are formally our assumptions:

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 $\gamma_j > 0$ 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 $\beta_1 + \gamma_1 < 1$, 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 $\gamma_j \geq 0$). It is needed in establishing the irreducibility and aperiodicity of the process. This is not really restrictive since if $\beta_j = \gamma_j = 0$ 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 or when the conditional variance was large. 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 β -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 β -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 exist different numerical techniques for solving this problem; Tong (1996) for an overview. It is important to note that the process is strictly stationary regardless of the parameter values in 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 (β_1 and γ_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 $E(\beta_1 + \gamma_1 u_t^2)^k < 1$ 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 existences of moment are given in Corollaries 1 and 3.

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

This corollary implies that it is not necessary that the covariance-stationarity requirement, $\beta_j + \gamma_j < 1$, 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 $\beta_j + \gamma_j$ 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 Paolella (2004a); see also Francq, Roussignol, and Zakoian (2001).

Corollary 2 *Under assumptions A1, A2, A4, and constant probabilities, if the GARCH processes (4) for $s_t > 1$ are integrated (IGARCH) and $p_{1t} > 0$, 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 volatility process are IGARCH in the non-stable regimes, the RS-GARCH process is covariance-stationary regardless of the features of the state probabilities if there exists a real number δ such that $0 < \delta \leq p_{1t}$. 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 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 $\beta_j + \gamma_j$ 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-A3, and constant probabilities, if $\sum_{j=1}^n p_j E(\beta_j + \gamma_j u_t^2)^k < 1$ for some $k \geq 1$, the results of Theorem 2 follow.*

This corollary shows that, in terms of existence of moments, the condition $E(\beta_j + \gamma_j u_t^2)^k < 1$ 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 and Prediction

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 variable. Indeed, the conditional variance σ_t^2 depends directly on the state variable s_t and σ_{t-1}^2 , which itself depends on s_{t-1} and σ_{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 numerically infeasible 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 such a difficulty. 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 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.

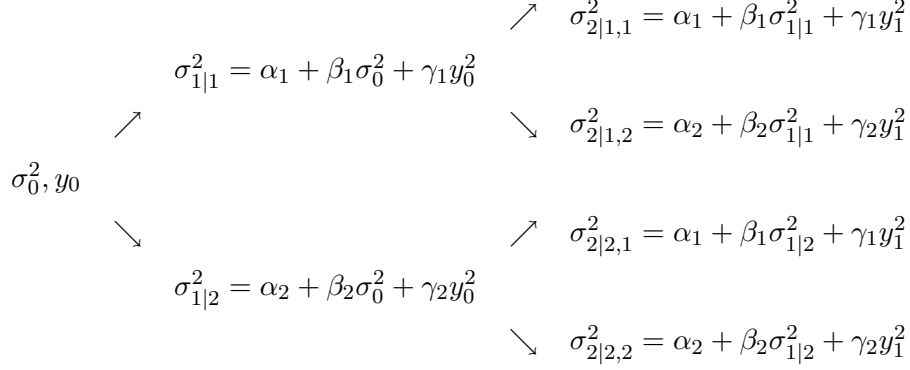


Figure 1: The problem of path dependence in a RS-GARCH model with two states

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, like a Student, 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), (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. In the current presentation, we change equation (4) to

$$\sigma_t^2 = \alpha_{s_t} + \beta_{s_t} \sigma_{t-1}^2 + \gamma_{s_t} y_{t-1}^2 \quad (7)$$

because it renders the algorithm quicker (see the remark in Section 3.1.3). We stress that there is no additional intrinsic difficulty in dealing with the specification (4) rather than (7). Moreover the theorems and corollaries of Section 2 also apply to this model although the proofs have to be adjusted.

The model parameters consist of $\delta = (\delta_0, \delta_1)'$, $\mu = (\mu_1, \mu_2)'$, and $\theta = (\theta'_1, \theta'_2)'$, where $\theta_k = (\alpha_k, \beta_k, \gamma_k)'$ for $k = 1, 2$. Given μ , θ , and the vector of states $S = (s_1, s_2, \dots, s_T)$, the

joint density of the observations $y = (y_1, y_2, \dots, y_T)$ is given by

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

where σ_t^2 is a function of θ through s_t , defined by equation (7). This would be the likelihood function to maximize if the states were known. Notice that it does not depend on δ . Given δ and y the distribution of S is given by

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

which does not depend on μ and θ . Given (8) and (9), we implement a Gibbs sampling algorithm that allows to sample from the full conditional posterior densities of blocks of parameters given by θ , μ , δ , and the elements of S . We define the prior densities of θ , μ , and δ when we explain the corresponding full conditional posterior densities in Sections 3.1.1-3.1.4. We explain how to compute predictive results in Section 3.1.5.

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 the problem of path dependence). Therefore,

$$\varphi(s_t | S_{\neq t}, \mu, \theta, \delta, y) \propto p_{1t}^{2-s_t} (1 - p_{1t})^{s_t-1} \prod_{j=t}^T \sigma_j^{-1} \exp\left(-\frac{(y_j - \mu_j)^2}{2\sigma_j^2}\right) \quad (10)$$

where $S_{\neq t}$ is the vector S without the element s_t , and μ_j and σ_j^2 depend on s_j since it defines the parameters operating at time j . We see that 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 (10) for $s_t = 1$ and 2 , normalize and draw as from a Bernoulli distribution.

3.1.2 Sampling θ

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

$$\varphi(\theta | S, \mu, \delta, y) \propto \pi(\theta) \prod_{t=1}^T \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_t)^2}{2\sigma_t^2}\right), \quad (11)$$

which does not depend on δ . We **sample θ 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 (11), compute $\kappa(\alpha_1 | S^{(r)}, \beta_1^{(r)}, \gamma_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$, the kernel of the conditional posterior density of α_1 given the values of S , β_1 , γ_1 , θ_2 , and μ sampled at iteration n , over a grid $(\alpha_1^1, \alpha_1^2, \dots, \alpha_1^G)$, 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

$$f_i = \int_{\alpha_1^1}^{\alpha_1^i} \kappa(\alpha_1 | S^{(r)}, \beta_1^{(r)}, \gamma_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y) d\alpha_1, \quad i=2, \dots, G. \quad (12)$$

3. Generate $u \sim U(0, f_G)$ and invert $f(\alpha_1 | S^{(r)}, \beta_1^{(r)}, \gamma_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$ by numerical interpolation to get a draw $\alpha_1^{(r+1)} \sim \varphi(\alpha_1 | S^{(r)}, \beta_1^{(r)}, \gamma_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$.

4. Repeat steps 1-3 for $\varphi(\beta_1 | S^{(r)}, \alpha_1^{(r+1)}, \gamma_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y)$,
 $\varphi(\gamma_1 | S^{(r)}, \alpha_1^{(r+1)}, \beta_1^{(r+1)}, \theta_2^{(r)}, \mu^{(r)}, y)$, $\varphi(\alpha_2 | S^{(r)}, \beta_2^{(r)}, \gamma_2^{(r)}, \theta_1^{(r+1)}, \mu^{(r)}, y)$, etc.

Note that intervals of values for the elements of θ_1 and θ_2 must be defined. The choice of these bounds (such as α_1^1 and α_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 μ

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

$$\varphi(\mu | S, \theta, \delta, y) \propto \pi(\mu) \prod_{t=1}^T \exp\left(-\frac{(y_t - \mu_t)^2}{2\sigma_t^2}\right) \quad (13)$$

which does not depend on δ . Given S , we can split the observations into one group belonging to regime 1, and the rest which belongs to regime 2, so that

$$\varphi(\mu | S, \theta, \delta, y) \propto \pi(\mu) \prod_{t|s_t=1} \exp\left(-\frac{(y_t - \mu_1)^2}{2\sigma_t^2}\right) \prod_{t|s_t=2} \exp\left(-\frac{(y_t - \mu_2)^2}{2\sigma_t^2}\right). \quad (14)$$

The factor $\prod_{t|s_t=1} \exp\left(-\frac{(y_t - \mu_1)^2}{2\sigma_t^2}\right)$ can be written as a normal density in μ_1 , with mean $\bar{\mu}_1$ and variance τ_1^2 defined by

$$\tau_1^2 = \left(\sum_{t|s_t=0} \sigma_t^{-2} \right)^{-1} \quad (15)$$

$$\bar{\mu}_1 = \tau_1^2 \sum_{t|s_t=0} \sigma_t^{-2} y_t. \quad (16)$$

The second factor $\prod_{t|s_t=2} \exp\left(-\frac{(y_t - \mu_2)^2}{2\sigma_t^2}\right)$ is also a normal density in μ_2 . If the prior factorizes as $\pi(\mu) = \pi(\mu_1)\pi(\mu_2)$, the posterior factorizes in the same way and we can sample μ_1 independently of μ_2 in this block. If $\pi(\mu_i)$ is a normal density (respectively uniform density), $\pi(\mu_i | S, \theta, \delta, y)$ is also normal (respectively truncated normal). If $\pi(\mu)$ is bivariate normal, with a non-zero correlation, the joint posterior is also bivariate normal and can be simulated easily.

Remark: this part of this algorithm gets heavier if we use equation (4) instead of equation (7). The reason is that we cannot express $\prod_{t|s_t=1} \exp\left(-\frac{(y_t - \mu_1)^2}{2\sigma_t^2}\right)$ as a normal density in μ_1 because μ_1 also appears in σ_t^2 . Moreover we have to put the factor σ_t^{-1} back in (13). Then we must sample μ_1 and μ_2 jointly and numerically (by griddy-Gibbs, or a Metropolis step).

3.1.4 Sampling δ

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

$$\varphi(\delta|S, \mu, \theta, y) \propto \pi(\delta) \prod_{t=1}^T p_{1t}^{2-s_t} (1 - p_{1t})^{s_t-1}, \quad (17)$$

which does not depend on μ and θ . We sample $\delta = (\delta_0, \delta_1)'$ with the griddy-Gibbs sampler, see Section 3.1.2, so that given prior intervals, the prior can be chosen as we wish, for example as uniform densities.

3.1.5 Predictive Results

The outcome of the Gibbs sampler described above is a sample of values of the parameters of the model, generated from the posterior distribution. We denote by Ψ the parameter vector, i.e. $\Psi = (\mu', \theta', \delta', S')$, and by $\Psi^{(r)}$, for $r = 1, 2, \dots, R$, the sample of size R obtained by running the Gibbs sampler. Given the latter, it is straightforward to compute by simulation predictive densities of future observations and features of these. In particular, the quantile of level a of a predictive distribution serves to compute the value-at-risk (VaR) of an investment, for the risk level a . Another feature of interest is the probability integral transforms of a series of future observations that are not used for estimation but are saved for density forecast evaluation of the model.

The predictive density of the first post-sample observation evaluated at the value y_{T+1} is defined as

$$f(y_{T+1}|y) = \int f(y_{T+1}|\Psi, y) \varphi(\Psi|y) d\Psi, \quad (18)$$

where $f(y_{T+1}|\Psi, y)$ is $N(\mu_1, \sigma_{T+1}^2(1))$ with probability $p_{1,T+1}$ and $N(\mu_2, \sigma_{T+1}^2(2))$ with probability $1 - p_{1,T+1}$, and $\sigma_T^2(j)$, for $j = 1$ or 2 , means that we evaluate σ_t^2 with the parameters of the j -th regime. Therefore, we can approximate $f(y_{T+1}|y)$ by

$$f(y_{T+1}|y) \approx \frac{1}{R} \sum_{r=1}^R f(y_{T+1}|\Psi^{(r)}, y). \quad (19)$$

If we repeat this evaluation for m different values of y_{T+1} that we choose, we can normalize the predictive density by applying a deterministic integration rule using the m points, and simultaneously compute any moment of this density. We can also compute any quantile or probability as a by-product of the deterministic integration.

A simple alternative procedure consists in generating R random draws of the predictive density as follows:

Step 1: simulate $\Psi^{(r)} \sim \varphi(\Psi | y)$, which is done by the Gibbs sampler.

Step 2: simulate $y_{T+1}^{(r)} \sim f(y_{T+1} | \Psi^{(r)})$. Go to step 1.

Using the sample $y_{T+1}^{(r)}$, $r = 1, 2, \dots, R$ of the predictive density, it is then straightforward to approximate the density by a kernel estimation and to approximate moments and other features of the density by the corresponding sample quantities. Although this simple alternative

method is in principle less precise than the procedure based on deterministic integration, it can be easily **generalized to multi-step prediction**. Extending (18), we have

$$\begin{aligned}
f(y_{T+s} | y) = & \int \left[\int \int \dots \int f(y_{T+s} | y_{T+s-1}, \dots, y_{T+1}, y, \Psi) \times \right. \\
& f(y_{T+s-1} | y_{T+s-2}, \dots, y_{T+1}, y, \Psi) \times \dots \\
& \times f(y_{T+1} | y, \Psi) dy_{T+s-1} dy_{T+s-2} \dots dy_{T+1} \left. \right] \\
& \varphi(\Psi | y) d\Psi,
\end{aligned} \tag{20}$$

for which draws can be obtained by extending the above algorithm to a $(s+1)$ -step algorithm. The draw of y_{T+1} serves as conditioning information to draw y_{T+2} , both realisations serve to draw y_{T+3} , etc. Each of these draws is generated from the corresponding finite mixture of two normal densities defined by the model. Note that to generate a draw from the mixture density of y_{T+j} , we compute the probability $p_{1,T+j}$ as defined in (6), using the simulated value of y_{T+j-1} as input, we draw s_{T+j} as a Bernoulli variable, and finally we draw y_{T+j} from the $N(\mu_{s_{T+j}}, \sigma_{T+j}^2(s_{T+j}))$ density.

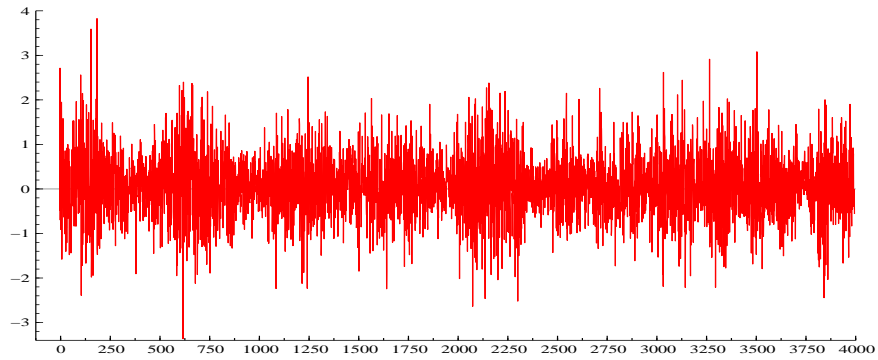
3.2 Simulation Example

We have simulated a DGP corresponding to the model defined by equations (3)-(7)-(6), with parameter values as reported in the second column of Table 2, and $u_t \sim N(0, 1)$. One component is stable ($\beta_1 + \gamma_1 = 0.95$) and the other is explosive ($\beta_2 + \gamma_2 = 1.40$). 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. Finally, the mean of the data is estimated to be 0.086, which is approximately equal to $0.92\mu_1 + (1 - 0.92)\mu_2$, with $\mu_1 = -\mu_2 = 0.10$. The density is symmetric, and its excess kurtosis, estimated to be 1.56, is much larger than 0.025, 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 $\rho_1 = 0.0238$ and decline according to $\rho_n = 0.95\rho_{n-1}$, so that $\rho_{50} = 0.0019$. The corresponding values in panel (c) of Figure 2 are 0.172 for ρ_1 and 0.037 for ρ_{50} . Thus, it is clear that, due to the explosive component, the ACF declines relatively slowly, although the decline rate is exponential.

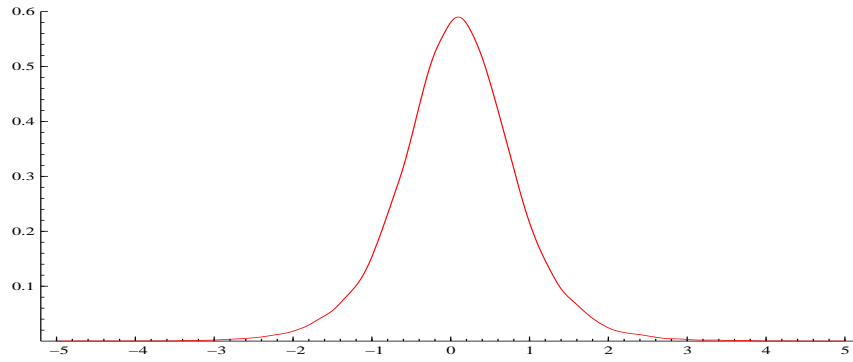
Table 1: Descriptive statistics for simulated data

Mean	0.086	Maximum	5.29
Standard deviation	0.779	Minimum	-4.85
Skewness	-0.015	Kurtosis	4.56

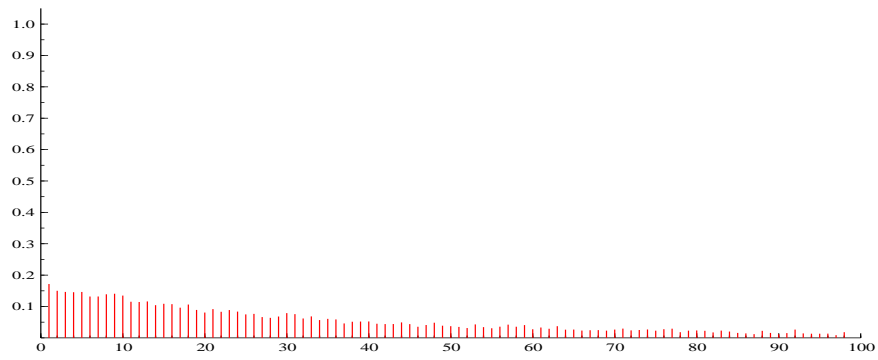
Statistics for 50,000 observations of the DGP defined in Table 2.



(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
α_1	0.0050 (0.0001,0.025)	0.0036 (0.0026)	0.0059 (0.0015)
β_1	0.930 (0.85,0.99)	0.937 (0.022)	0.931 (0.0084)
γ_1	0.0200 (0.001,0.08)	0.0193 (0.012)	0.0584 (0.0070)
α_2	0.100 (0.01,0.30)	0.125 (0.073)	-
β_2	1.200 (0.75,1.60)	1.201 (0.15)	-
γ_2	0.200 (0.02,0.60)	0.160 (0.13)	-
μ_1	0.100 -	0.070 (0.010)	0.0723 (0.0071)
μ_2	-0.100 -	0.076 (0.072)	-
δ_0	-2.40 (-3.50,-1.60)	-2.52 (0.47)	-
δ_1	-0.200 (-0.80,0)	-0.496 (0.21)	-

Posterior standard deviations under posterior means.
Sample of 4,000 observations from DGP.
DGP: equations (3)-(7)-(6) with $N(0, 1)$ distribution.
RS-GARCH: same model as DGP.
 N -GARCH: equations (1)-(2), and $N(0, 1)$ distribution.

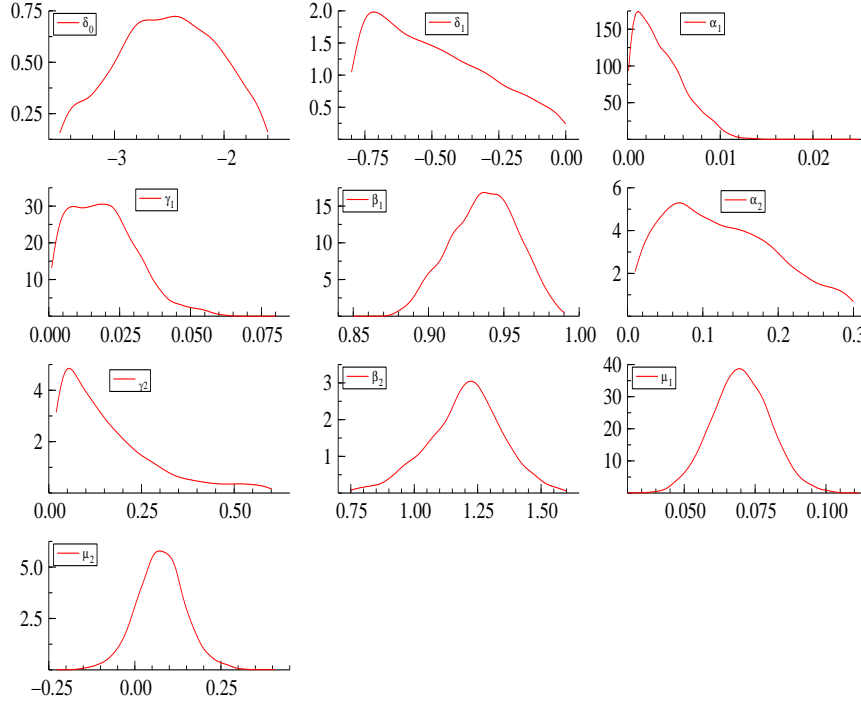


Figure 3: Posterior densities for the RS-GARCH model (simulated DGP)

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. In Figure 3, we report the corresponding posterior densities. 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 griddy-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. They show that the posterior means are close to the true values for the sample we use (4,000 observations), with two exceptions: for μ_2 the true value is 2.5 standard deviations away from the posterior mean, and for δ_1 , it is 1.5 times away. 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 come from the second regime. The computing time for generating 10,000 draws was about 210 hours using a 2.6 Ghz Intel Xeon processor.

The last column of Table 2 contains the results for a GARCH(1,1) model with $N(0,1)$ 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. These results illustrate that the posterior mean of the persistence effect ($\gamma_1 + \beta_1$), equal to 0.99, is pushed towards 1 due to the specification error. 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 4 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), (7), and (6), the same model with a fixed probability, 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 ν degrees of freedom), and the MN-GARCH(2,2) model of Haas et al. (2004), defined as

$$y_t = \begin{cases} \mu_1 + \sigma_{1,t}u_t, & \text{with probability } 1 - p \\ \mu_2 + \sigma_{2,t}u_t, & \text{with probability } p, \end{cases} \quad (21)$$

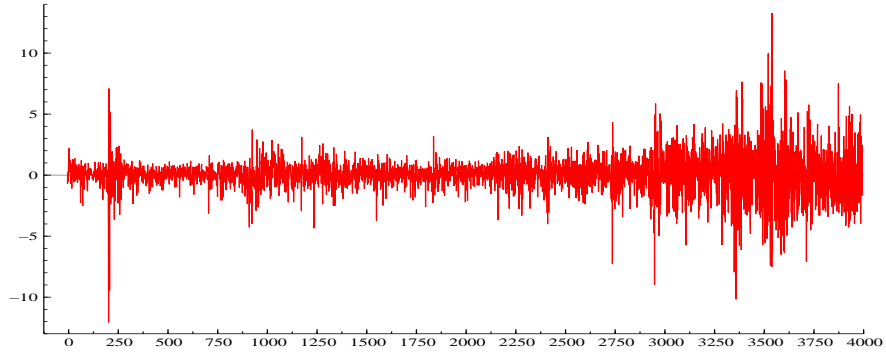
$$\sigma_{1,t}^2 = \alpha_1 + \beta_1\sigma_{1,t-1}^2 + \gamma_1y_{t-1}^2, \quad (22)$$

$$\sigma_{2,t}^2 = \alpha_2 + \beta_2\sigma_{1,t-1}^2 + \gamma_2y_{t-1}^2, \quad (23)$$

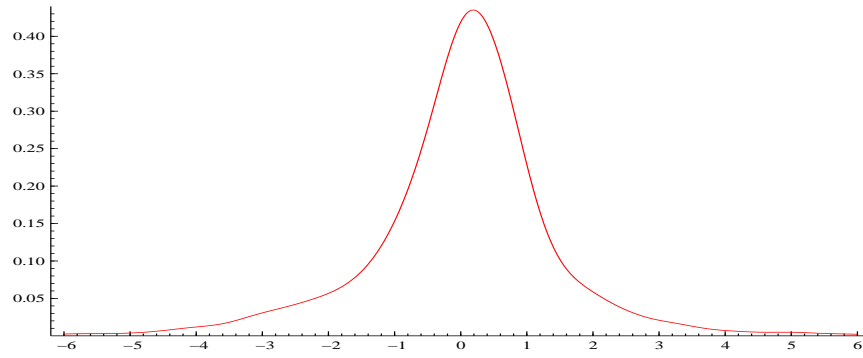
$$u_t \sim i.i.d. N(0, 1). \quad (24)$$

Equivalently, the distribution of y_t is the Gaussian mixture $(1 - p)N(\mu_1, \sigma_{1,t}^2) + pN(\mu_2, \sigma_{2,t}^2)$. For identification we impose that p is larger than 0.5. Bauwens and Rombouts (2005) explain how to estimate this model by Bayesian inference. However, we do not impose that the mixture mean is equal to zero, hence μ_1 and μ_2 are not linked by imposing that $p\mu_1 + (1 - p)\mu_2$ is equal to 0. Note that for the models with a fixed probability (RS-GARCH-2 and MN-GARCH), we report the results for $p = (1 + \exp(\delta_0))^{-1}$ rather than δ_0 . Some comments about the estimation results follow.

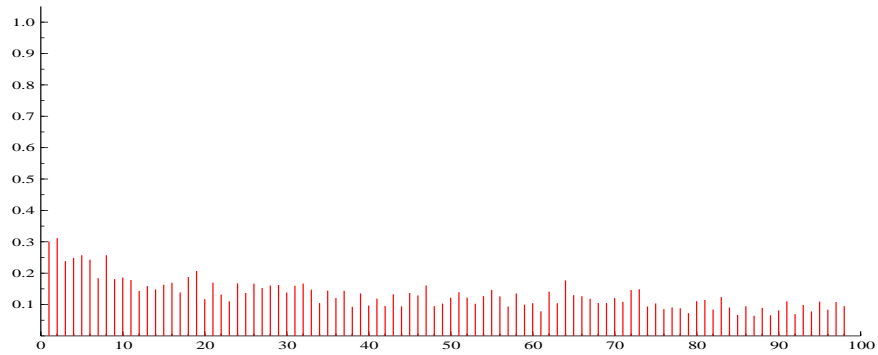
1) RS-GARCH-1: the marginal posterior densities are shown in Figure 5. The intervals over which the densities are drawn are the prior intervals (except for μ_1 and μ_2 since the prior is not truncated for these parameters). These intervals were chosen to avoid negative values, and by trial and error so as to avoid truncation, though it still occurs for α_2 . The density of α_2 is apparently bimodal. The first regime is stable, with the posterior mean of $\beta_1 + \gamma_1$ equal to 0.93 and the second one is explosive, with the posterior mean of $\beta_2 + \gamma_2$ equal to 1.92. There is not much evidence for a time-varying probability, since the posterior mean of δ_1 is equal to 0.037 with standard deviation equal to 0.025. Computing the time-varying probability of the first regime at the posterior mean of δ_0 and δ_1 , we get the value 0.91 if



(a) Sample path (4,000 observations)



(b) Kernel density



(c) Correlogram of squared data

Figure 4: NASDAQ daily returns from 23/12/1986 to 28/10/2002

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

	RS-GARCH-1	RS-GARCH-2	MN-GARCH	N -GARCH	t -GARCH
α_1	0.0063 (0.0035)	0.0071 (0.0037)	0.0008 (0.0007)	0.0217 (0.0033)	0.0073 (0.0028)
β_1	0.924 (0.023)	0.915 (0.029)	0.922 (0.010)	0.848 (0.012)	0.887 (0.020)
γ_1	0.0058 (0.0049)	0.0068 (0.0058)	0.057 (0.0087)	0.148 (0.013)	0.082 (0.016)
α_2	0.089 (0.051)	0.104 (0.066)	0.121 (0.049)	-	-
β_2	1.499 (0.24)	1.457 (0.25)	0.779 (0.157)	-	-
γ_2	0.420 (0.195)	0.315 (0.201)	0.631 (0.055)	-	-
μ_1	0.106 (0.015)	0.106 (0.016)	0.165 (0.018)	0.082 (0.0098)	0.105 (0.013)
μ_2	-0.077 (0.088)	-0.068 (0.088)	-0.641 (0.118)	-	-
δ_0/p	$\delta_0 = -2.36$ (0.34)	$p = 0.895$ (0.052)	$p = 0.845$ (0.029)	-	-
δ_1	-0.037 (0.0254)	-	-	-	-
ν	-	-	-	-	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-1: equations (3)-(7)-(6) (time-varying probability).

RS-GARCH-2: (3)-(7)-(6) with $\delta_1 = 0$ (fixed probability p).

MN-GARCH: (21)-(24).

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

t -GARCH: equations (1)-(2), and t_ν distribution.

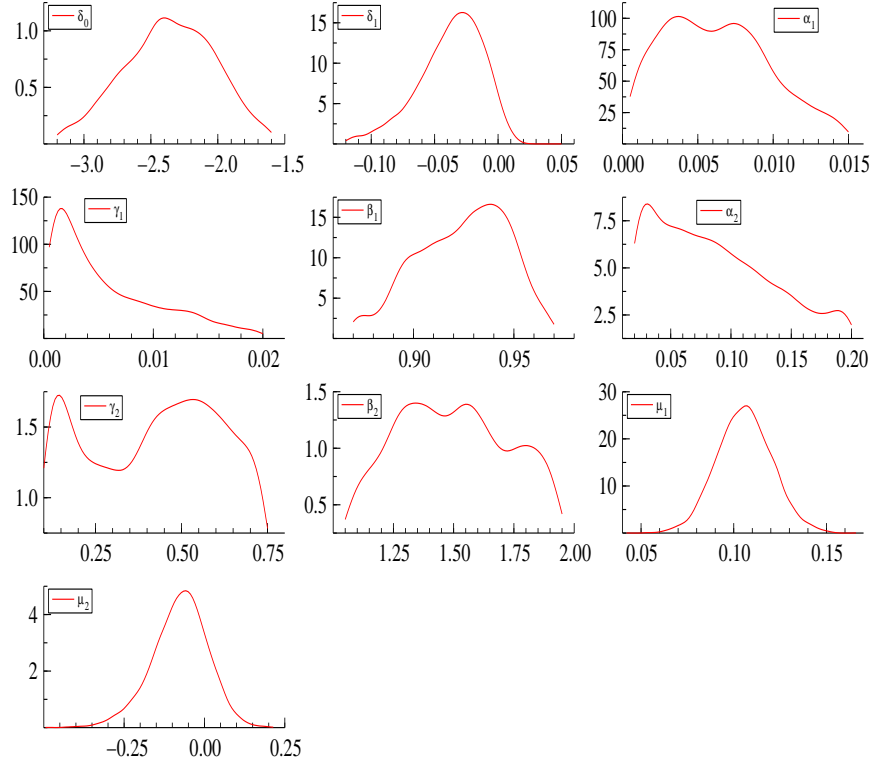


Figure 5: Posterior densities for the RS-GARCH-1 model (NASDAQ daily returns)

$y_{t-1} = 0$ and 0.94 if $|y_{t-1}| = 13.26$ (the largest observed absolute return). We notice that the posterior mean of μ_2 has a large proportion of its mass on negative values and its mean is -0.077 . Actually, the second regime is mostly associated with large negative returns, which explains the negative estimate of μ_2 , and is capturing the leverage effect.

2) RS-GARCH-2: the results are close to those for the time-varying probability model, which is not a surprise given the previous comments. The posterior mean of the fixed probability p of the first regime is about 0.90, which is marginally smaller than the average value of the time-varying probability in the first model. However, we notice that the posterior density of p is slightly bimodal, with a secondary mode at 0.82 and a long left tail. The marginal posterior densities are shown in Figure 6.

3) MN-GARCH: the marginal posterior densities are shown in Figure 7. 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, smaller than for the RS-GARCH model with fixed probability. The GARCH parameters of both models are either very different (α_1 , γ_1 , γ_2 , β_2) or very similar (β_1 , α_2), and the sum $\gamma_2 + \beta_2$ is estimated to be equal to 1.77 in the RS-GARCH-2 model and 1.41 in the MN-GARCH one. The weak stationarity condition for this model, see Haas et al. (2004), is that $[(1 - \beta_1 - \gamma_1)p / (1 - \beta_1) + (1 - \beta_2 - \gamma_2)(1 - p) / (1 - \beta_2)](1 - \beta_1)(1 - \beta_2)$ is positive. Evaluated at

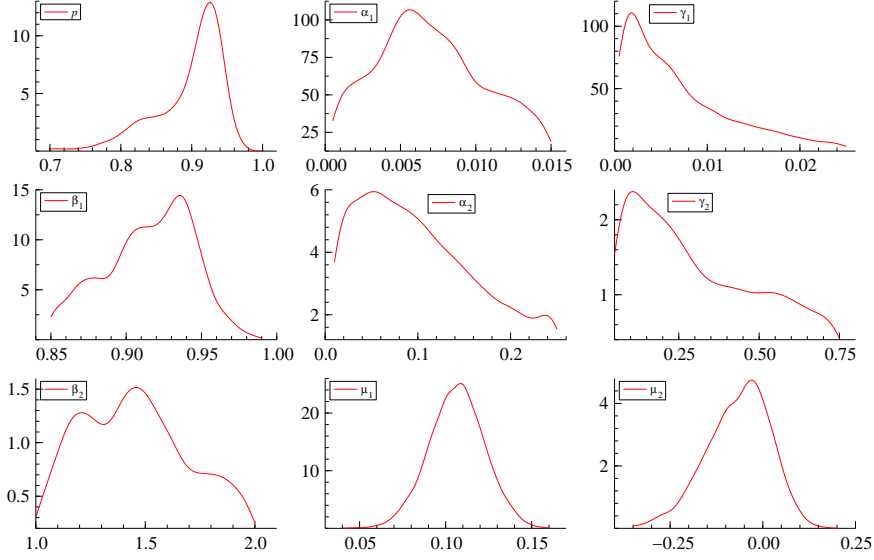


Figure 6: Posterior densities for the RS-GARCH-2 model (NASDAQ daily returns)

the posterior mean, this is equal to -0.0011 , but the mean of these quantities over the Gibbs draws is equal to 0.0025 (with a large variance due to some extreme values).

4) Normal-GARCH: the posterior mean of the persistence parameter ($\beta_1 + \gamma_1$) is equal to 0.996 , which makes it almost an IGARCH model. The model with a single component is not enough flexible.

5) t -GARCH: the posterior mean of the degrees of freedom parameter is close to 7 . The relevant persistence parameter for this model is $\beta_1 + \gamma_1 \nu / (\nu - 2)$. Evaluated at the posterior means, this is equal to 1.002 indicating again that the process is integrated.

5 Conclusion

We have developed a regime-switching univariate GARCH model with a time-varying probability of switching between a non-explosive regime and an explosive one. 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 Grays 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 proposed with the **approximate ML solutions**. Very preliminary results tend to indicate that the approximate estimators are strongly biased. Another idea to improve upon these approximate ML estimators is to use indirect inference since the RS-GARCH model can be easily simulated.

More research is needed about the regime switching GARCH model. Current and future work involves:

- Estimating the model with equation (4) rather than (7).

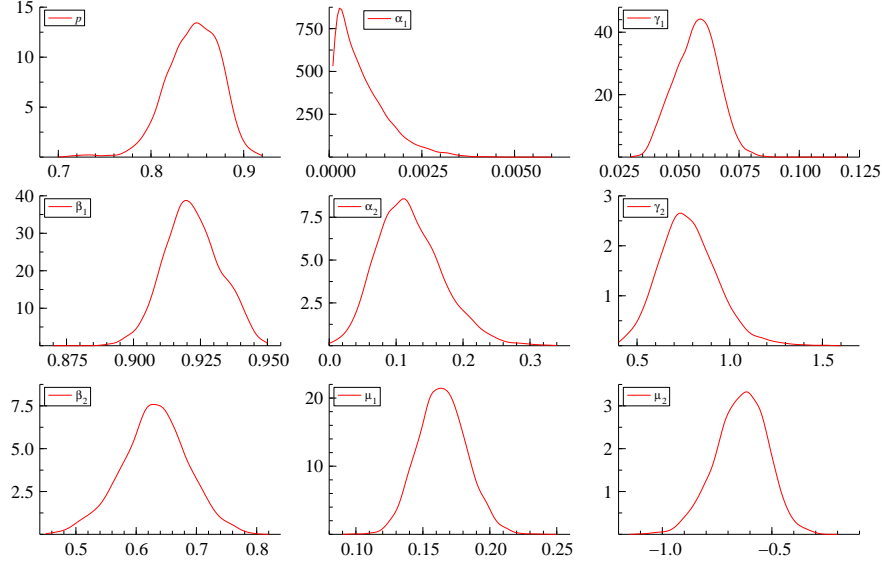


Figure 7: Posterior densities for the MN-GARCH model (NASDAQ daily returns)

- In-sample and out-sample comparisons of the RS-GARCH models with other GARCH models (such as those estimated in this paper).
- Estimating the model with other data series.
- Making the Bayesian algorithm faster, e.g. we can try to use a Metropolis algorithm rather than the griddy-Gibbs for the GARCH parameters.

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 σ_t^2 is a function of information dated at $t - 1$ or earlier, not of information dated at t . Then we write

$$y_t = \mu_j + \sqrt{h_{t-1}}u_t, \text{ with probability } p_j(y_{t-1}^2) \quad (25)$$

$$h_t = \alpha_j + \beta_j h_{t-1} + \gamma_j \epsilon_t^2 = \alpha_j + (\beta_j + \gamma_j u_t^2)h_{t-1} \quad (26)$$

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 $(\mathbb{R} \times \mathbb{R}_+, \mathfrak{S}, \lambda)$, where $\mathbb{R}_+ = (0, \infty)$, \mathfrak{S} is the Borel σ -algebra on $\mathbb{R} \times \mathbb{R}_+$, and λ denotes the usual Lebesgue measure on $(\mathbb{R} \times \mathbb{R}_+, \mathfrak{S})$. Let $P^m((y, h), A) = P((y_t, h_t) \in A | y_{t-m} = y, h_{t-m} = h)$ denote the probability that $\{y_t, h_t\}$ moves from (y, h) to the set $A \in \mathfrak{S}$ in m steps. In order to establish the geometric ergodicity of the Markov chain, we first show that the process is aperiodic and λ -irreducible. To prove these features it is sufficient to show that $P^m((y_0, h_0), A) > 0$ for some $m \geq 1$, for all $(y_0, h_0) \in \mathbb{R} \times \mathbb{R}_+$ and all Borel measurable

sets $A \in \mathfrak{S}$ with positive Lebsgue measure (Chan (1993)). By solving recursively the state space equations we can show that for a given (y_0, h_0) we can reach any $(y, h) \in A$ through the following steps (note that h_t is determined given y_t): $(h_0, y_0) \rightarrow (h_0, \tilde{y}_{ij\ell}) \rightarrow (\tilde{h}_{ij}, \tilde{y}_{ij\ell}) \rightarrow (\tilde{h}_{ij}, y) \rightarrow (h, y)$, that is for the set $A \in \mathfrak{S}$, we can define the following ‘connecting’ sets:

$$A_{ij\ell} = \left\{ ((\tilde{y}_{ij\ell}, y, \tilde{h}_{ij}) \in \mathbb{R}^2 \times \mathbb{R}_+ : \right. \\ \left. (h, y) \in A, \tilde{h}_{ij} = \frac{[h - \alpha_j - \gamma_j(y - \mu_i)^2]}{\beta_j}, (\tilde{y}_{ij\ell} - \mu_\ell)^2 = \frac{\tilde{h}_{ij} - \alpha_i - \beta_i h_0}{\gamma_i} \right\}$$

where $i, j, \ell \in \{1, \dots, n\}$. The continuity of the state space equations implies that $\lambda(A_{ij\ell}) > 0$ and given assumption A3 there exists a real number δ such that $0 < \delta \leq p_1(\cdot)$, hence

$$\begin{aligned} P^2((y_0, h_0), A) &\geq \sum_{(i,j,\ell) \in \{1,\dots,n\}_{A_{ij\ell}}} \int p_\ell(y_0^2) p_i(\tilde{y}_{ij\ell}^2) p_j(y^2) f_{ij\ell} \left(\frac{\tilde{y}_{ij\ell} - \mu_\ell}{\sqrt{h_0}} \right) f_{ij} \left(\frac{y - \mu_i}{\sqrt{\tilde{h}_{ij}}} \right) d\tilde{y} dy dh \\ &\geq \int_{A_{111}} p_1(y_0^2) p_1(\tilde{y}_{111}^2) p_1(y^2) f_{111}(\cdot) f_{11}(\cdot) d\tilde{y} dy dh \\ &\geq \delta^3 \int_{A_{111}} f_{111}(\cdot) f_{11}(\cdot) d\tilde{y} dy dh > 0. \end{aligned}$$

It is also straightforward to show that for all $m \geq 2$, $P^m((y_0, h_0), A) > 0$, for all (y_0, h_0) . Hence the chain is λ -irreducible and aperiodic. We note that every λ -non null compact set is small and petite (Chan (1993)) and can serve as test set. Given these results, it remains to check the drift criterion of theorem 15.0.1 of Meyn and Tweedie (1993) to obtain the desired result. As drift function we use $V(y, h) = 1 + \bar{\eta}y^2 + h$, $\bar{\eta} = \eta - \beta_1 - \gamma_1$ where η is some positive number that satisfies $\beta_1 + \gamma_1 < \eta < 1$ and as compact set we use $C = \{(y, h) \in \mathbb{R} \times \mathbb{R}_+ : h + y^2 \leq c\}$, where c is a positive number to be determined below. Thus,

$$\begin{aligned} &\mathbb{E}[V(y_t, h_t) | h_{t-1} = h, y_{t-1} = y] = 1 + \bar{\eta}h + M + (\beta_1 + \gamma_1)h \\ &+ h \sum_{i=1}^n p_i(y^2) \sum_{j \neq 1} \mathbb{E} \left[p_j \left((\mu_i + \sqrt{h_t} u_t)^2 \right) h [(\beta_j + \gamma_j u_t^2) - (\beta_1 + \gamma_1 u_t^2)] \right] \\ &= 1 + h \left\{ \eta + \frac{M}{h} + \sum_{i=1}^n p_i(y^2) \sum_{j \neq 1} \mathbb{E} \left[p_j \left((\mu_i + \sqrt{h_t} u_t)^2 \right) [(\beta_j + \gamma_j u_t^2) - (\beta_1 + \gamma_1 u_t^2)] \right] \right\} \end{aligned}$$

where $M = \sum_{i,j} \left\{ \bar{\eta} \mu_i^2 p_i(y^2) + p_i(y^2) \mathbb{E} \left[\alpha_j p_j \left((\mu_i + \sqrt{h} u_t)^2 \right) \right] \right\}$. Assumption A3 implies that

$$\frac{M}{h} + \sum_{i=1}^n p_i(y^2) \sum_{j \neq 1} \mathbb{E} \left[p_j \left((\mu_i + \sqrt{h} u_t)^2 \right) [(\beta_j + \gamma_j u_t^2) - (\beta_1 + \gamma_1 u_t^2)] \right] \rightarrow 0 \text{ as } h \rightarrow \infty.$$

Since $\mathbb{E}[V(y_t, h_t) | h_{t-1} = h, y_{t-1} = y]$ is continuous on a compact set, it is bounded, and since $h < V(y, h)$, we can choose c such that $\mathbb{E}[V(y_t, h_t) | h_{t-1} = h, y_{t-1} = y] \leq \eta V(y, h) + a \mathbb{1}_C(y, h)$ (where $\mathbb{1}_C$ denotes an indicator function) for some $a < \infty$ and for all (y, h) , so that the drift criterion is satisfied. Hence the Markov chain $\{y_t, h_t\}$ is geometrically ergodic, and so is $\{y_t\}$.

The chain is also V-geometrically mixing (Meyn and Tweedie (1993), theorem 16.0.1), which is equivalent to the requirement that the process is β -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 $\Delta = E(u_t^{2k})$, $\bar{\eta} = \eta - E(\beta_1 + \gamma_1 u_t^2)^k$, and η is a positive number that satisfies $E(\beta_1 + \gamma_1 u_t^2)^k < \eta < 1$. The relevant compact set is $C = \{(y, h) \in \mathbb{R}_+ \times \mathbb{R} : h + y^2 \leq c\}$. After some tedious calculations, we have

$$\begin{aligned} & E[V(y_t, h_t)|h_{t-1} = h, y_{t-1} = y] = 1 + \bar{\eta}h^k + O(h^{k-1}) E(\beta_1 + \gamma_1 \epsilon_t)^k h^k \\ & + \sum_{i=1}^n p_i(y^2) \left\{ O(h^{k-1}) + h^k \sum_{j \neq 1} E \left[p_j((\mu_i + \sqrt{h}u_t)^2) [(\beta_j + \gamma_j u_t^2)^k - (\beta_1 + \gamma_1 u_t^2)^k] \right] \right\} \\ & = 1 + h^k \left\{ \eta + \frac{O(h^{k-1})}{h^k} + \sum_{i=1}^n p_i(y^2) \sum_{j \neq 1} E \left[p_j((\mu_i + \sqrt{h}u_t)^2) [(\beta_j + \gamma_j u_t^2)^k - (\beta_1 + \gamma_1 u_t^2)^k] \right] \right\}. \end{aligned}$$

Assumption A2 and the boundedness of the moment of the error term imply that

$$\frac{O(h^{k-1})}{h^k} + \sum_{i=1}^n p_i(y^2) \sum_{j \neq 1} E \left[p_j((\mu_i + \sqrt{h}u_t)^2) [(\beta_j + \gamma_j u_t^2)^k - (\beta_1 + \gamma_1 u_t^2)^k] \right] \rightarrow 0 \text{ as } h \rightarrow \infty.$$

Using the same arguments as in Theorem 1, we can show that there exists a compact set C such that $E[V(y_t, h_t)|h_{t-1} = h, y_{t-1} = y] \leq \eta V(y, h) + a 1_C(x)$ for some $a < \infty$ and for all (y, h) . Theorem 14.3.7 of Meyn and Tweedie (1993) then implies the desired result.

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

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 $\bar{\eta} = \eta - \sum_{j=1}^n p_j E(\beta_j + \gamma_j u_t^2)^k$ 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.
- BAUWENS, L., AND J. ROMBOUTS (2005): “Bayesian inference for the mixed conditional heteroskedasticity model,” CORE Discussion Paper 2005/85.
- 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.
- 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.
- DUEKER, M. (1997): "Markov Switching in GARCH Processes in Mean Reverting Stock Market Volatility," *Journal of Business and Economics Statistics*, 15, 26–34.
- 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.