Theory and Inference for a Markov Switching Garch Model

L. Bauwens, A. Preminger and J. Rombouts

Département des Sciences Économiques
de l'Université catholique de Louvain

UCL

THEORY AND INFERENCE FOR A
MARKOV SWITCHING GARCH MODEL

Luc Bauwens[1], Arie Preminger,[2] and Jeroen V.K. Rombouts[3]

August 13, 2007

Abstract

We develop a Markov-switching GARCH model (MS-GARCH) wherein the conditional mean and variance switch in time from one GARCH process to another. The switching is governed by a hidden Markov chain. We provide sufficient conditions for geometric ergodicity and existence of moments of the process. 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 illustrate the model on SP500 daily returns.

Keywords: GARCH, Markov-switching, Bayesian inference.

JEL Classification: C11, C22, C52.

---

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

[2]Department of Economics, University of Haifa, 31905 Israel.

[3]Institute of Applied Economics at HEC Montréal, CIRANO, CIRPEE, CREF, CORE (Université catholique de Louvain), CREF, 3000 Côte Sainte Catherine, H3T 2A7 Montréal (QC), Canada.

1 Introduction

The volatility of financial markets has been the object of numerous developments and applications over the past two decades, both theoretically and empirically. In this respect, the most widely used class of models is certainly that of GARCH models (see e.g. Bollerslev, Engle, and Nelson (1994) for an overview). These models usually indicate a high persistence of the conditional variance (i.e. a nearly integrated GARCH process). Diebold (1986) and Lamoureux and Lastrapes (1990), among others, argue that the nearly integrated behavior of the conditional variance may originate from structural changes in the variance process which are not accounted for by standard GARCH models. Furthermore, Mikosch and Starica (2004) show that estimating a GARCH model on a sample displaying structural changes in the unconditional variance does indeed create an integrated GARCH effect. These findings clearly indicate a potential source of misspecification, to the extent that the form of the conditional variance is relatively inflexible and held fixed throughout the entire sample period. Hence the estimates of a GARCH model may 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.

Indeed, several models based on the idea of regime changes have been proposed. Schwert (1989) consider a model in which returns can have a high or 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 Markov-switching parameters in order to take into account sudden changes in the level of the conditional variance. They use an ARCH specification instead of a GARCH to avoid the problem of path dependence of the conditional variance which renders the computation of the likelihood function infeasible. This occurs because the conditional variance at time t depends on the entire sequence of regimes up to time t due to the recursive nature of the GARCH process. Since the regimes are unobservable, one needs to integrate over all possible regime paths when computing the sample likelihood, but the number of possible paths grows exponentially with t, which renders ML estimation intractable. Gray (1996) presents a tractable Markov-switching GARCH model and a modification of his model is suggested by Klaassen (2002); see also Bollen, Gray, and Whaley (2000), Dueker (1997), Haas, Mittnik, and Paolella (2004), and Marcucci (2005) for related papers. Stationarity conditions for some of these tractable models are given by

Abramson and Cohen (2007).

The objective of this paper is to develop both the probabilistic properties and the estimation of a Markov swtiching GARCH (MS-GARCH) model that has a finite number of regimes in each of which the conditional mean is constant and the conditional variance takes the form of a GARCH(1,1) process. We provide sufficient conditions for the geometric ergodicity and the existence of moments of the proposed model. We find that for strict stationarity, it is not necessary that the stability condition of Nelson (1990) be satisfied in all the GARCH regimes but it must be satisfied on average with respect to the unconditional probabilities of the regimes. Further, for covariance stationarity, the GARCH parameters in some regimes can be integrated or even explosive.

Concerning the estimation method, we propose a Bayesian Markov chain Monte Carlo (MCMC) algorithm that circumvents the problem of path dependence by including the state variables in the parameter space and simulating them by Gibbs sampling. We illustrate by a simulation experiment that the algorithm is able to recover the parameters of the data generating process, and we apply the algorithm to a real data set. For the more simple MS-ARCH case, Francq, Roussignol, and Zakoian (2001) establish the consistency of the ML estimator. Bayesian estimation of a Markov switching ARCH model where only the constant in the ARCH equation can switch, as in Cai (1994), has been studied and illustrated by Kaufman and Fruhwirth-Schnatter (2002) and Kaufman and Scheicher (2006). Das and Yoo (2004) propose an MCMC algorithm for the same model (switch in the constant only) but with a GARCH term and therefore tackle the path dependence problem, but they do not provide an application to a real data series. Finally, the most comparable work to our paper (for estimation) is that of Henneke, Rachev, and Fabozzi (2006) who estimate by a MCMC algorithm a Markov-switching ARMA-GARCH model. They apply their algorithm to the data used by Hamilton and Susmel (1994). Non-Bayesian estimation of MS-GARCH models is studied by Francq and Zakoian (2005) who propose to estimate the model by the generalized method of moments.

This sequel of the paper is organized as follows: in Section 2, we define our version of the MS-GARCH model and state sufficient conditions for geometric ergodicity and existence of moments. In Section 3, we explain how the model can be estimated in the Bayesian framework and provide a numerical example. In Section 4, we apply our approach to financial data. and

in Section 5 we conclude and discuss possible extensions. Proofs of the theorems stated in the paper are gathered in an appendix.

2 Markov-Switching GARCH Model

The GARCH(1,1) model can be defined by

$$y_t = \mu_t + \sigma_t u_t \tag{1}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2}$$

where σ_t and μ_t are measurable functions of $y_{t-\tau}$ for $\tau \leq t - 1$, $\epsilon_t = y_t - \mu_t$, and the error term u_t is i.i.d. with zero mean and unit variance. In order to ensure easily the positivity of the conditional variance we impose the restrictions $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$. For simplicity, we assume that μ_t is constant. The sum $\alpha + \beta$ measures the persistence of a shock to the conditional variance in equation (2). When a GARCH 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 the second moment of the return process may not exist. 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 our idea to estimate a model that permits regime switching in the parameters. We call it a Markov-switching GARCH (MS-GARCH) model. It is a generalization of the GARCH model and permits a different persistence in the conditional variance of each regime. Thus, the conditional variance in each regime accommodates volatility clustering, nesting the GARCH model as special case. Let $\{s_t\}$ be an ergodic Markov chain on a finite set $S = \{1, \ldots, n\}$, with transition probabilities $\{\eta_{ij} = \mathrm{P}(s_t = i | s_{t-1} = j)\}$ and invariant probability measure $\{\pi_i\}$. The MS-GARCH model is given by

$$y_t = \mu_{s_t} + \sigma_t u_t \tag{3}$$

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

where $\omega_{s_t} > 0$, $\alpha_{s_t} \geq 0$, $\beta_{s_t} \geq 0$ for $s_t \in \{1, 2, \ldots n\}$, and $\epsilon_t = y_t - \mu_{s_t}$. These assumptions on the GARCH coefficients entail that σ_t^2 is almost surely strictly positive. Conditions for the weak stationarity and existence of moments of any order for the Markov-switching

GARCH(p,q) model with zero means μ_{s_t}, in which the discrete Markov chain of the latent states is initiated from its stationary probabilities, have been derived by Francq and Zakoian (2005), see also Francq, Roussignol, and Zakoian (2001). In related papers, Yang (2000), Yao and Attali (2000), Yao (2001), and Francq and Zakoian (2002) derived conditions for the asymptotic stationarity of some AR and ARMA models with Markov-switching regimes. The MS-GARCH process is not a Markov chain in general. However, the extended process $Z_t = (y_t, h_t, s_t)'$ is a Markov chain (see the Appendix). In what follows, we state mild regularity conditions for which this chain is geometrically ergodic and has finite moments. These results are based on Markov chain theory, see e.g. Meyn and Tweedie (1993) and Chan (1993). We impose the following assumptions:

A1 The error term u_t is i.i.d. with a continuous density on the whole real line which is centered on zero. Furthermore, $\mathrm{E}(|u_t^2|^\delta) < \infty$ for some $\delta > 0$.

A2 $\alpha_i > 0$, $\beta_i > 0$ and $\eta_{ii} \in (0, 1)$ for $i \in \{1, \ldots, n\}$.

A3 $\sum_{i=1}^{n} \pi_i \mathrm{E}[\log(\alpha_i u_t^2 + \beta_i)] < 0$.

The first assumption is satisfied for a wide range of distributions for the error term, e.g. the normal and the Student distributions. For $\delta = 1$, we set the variance to unity and if $\delta < 1$, the parameters of the conditional scaling factor of the data are estimated. The second assumption is slightly stronger than the non-negativity conditions of Bollerslev (1986) for the GARCH(1,1) model. Under this assumption all the regimes are accessible and the discrete Markov chain is ergodic. These assumptions are needed in order to establish the irreducibility and aperiodicity of the process. Assumption A3 implies that at least one of the regimes is stable. We assume, without loss of generality throughout that in the first regime ($s_t = 1$) the process is strictly stationary, thus $\mathrm{E}\log(\alpha_1 u_t^2 + \beta_1) < 0$. To obtain the results in Theorem 1, we observe that it is not necessary that the strict stationarity requirement of Nelson (1990) be satisfied for all the GARCH regimes but on average with respect to the invariant probability distribution of the latent states.

Theorem 1 *Under assumptions A1-A3, y_t is geometrically ergodic and if it is initiated from its stationary distribution, then the process is strictly stationary and β-mixing (absolutely regular) with exponential decay. Moreover, $\mathrm{E}(|y_t|^{2p}) < \infty$ for some $p \in (0, \delta]$ where the expectations are taken under the stationary distribution.*

4

The geometric ergodicity ensures not only that a unique stationary probability measure for the process exists, but also that the chain, irrespective of its initialization, converges to it at a geometric rate with respect to the total variation norm. Markov chains with this property satisfy conventional limit theorems such as the law of large numbers and the central limit theorem for any given starting value given the existence of suitable moments, see Meyn and Tweedie (1993, ch.17) for details. The definition of regular mixing can be found in Doukhan (1994, Section 1.1) who also shows that the exponential decaying rate of the mixing numbers implies that the autocovariance function converges to zero at least at the same rate. In order to establish the existence of higher order moments, we define the $n \times n$ matrix

$$\Omega = \begin{pmatrix} \mathrm{E}(\alpha_1 u_t^2 + \beta_1)^k \eta_{11} & \cdots & \mathrm{E}(\alpha_n u_t^2 + \beta_n)^k \eta_{n1} \\ \vdots & \ddots & \vdots \\ \mathrm{E}(\alpha_1 u_t^2 + \beta_1)^k \eta_{1n} & \cdots & \mathrm{E}(\alpha_n u_t^2 + \beta_n)^k \eta_{nn} \end{pmatrix}.$$

A similar matrix was first introduced by Yao and Attali (2000) for nonlinear autoregressive models with Markov switching. Let $\rho(\cdot)$ denote the spectral radius of a matrix, i.e. its largest eigenvalue in modulus. Then, we impose the following conditions:

A4 $\mathrm{E}(|u_t^2|^k) < \infty$ for some integer $k \geq 1$.

A5 $\rho(\Omega) < 1$.

Theorem 2 *Under assumptions A1-A2 and A4-A5, the process is geometrically ergodic and* $E(|y_t^2|^k) < \infty$ *for some integer* $k \geq 1$, *where the expectations are taken under the stationary distribution.*

The spectral radius condition used in Theorem 2 is simple to check in the leading case where $k = 1$. Let $d_i = \alpha_i + \beta_i$: if $d_i < 1$, assumption A5 is satisfied for this case, since $\eta_{ij} \in (0,1)$, see Lutkepohl (1996, p. 141, 4(b)), and the resulting MS-GARCH process is covariance-stationary. However, it is not necessary that all the GARCH processes of each regime be covariance-stationary. To illustrate this, we plot in Figure 1 the boundary curve $\rho(\Omega) = 1$ for $n = 2$ where $\eta_{11} = \eta_{22} = 0.85$. The covariance-stationarity region is the interior intersection of the boundary curve and the two axes. We observe that one of the GARCH

regimes does not need to be weakly stationary and can even be mildly explosive, provided that the other regime is sufficiently stable.
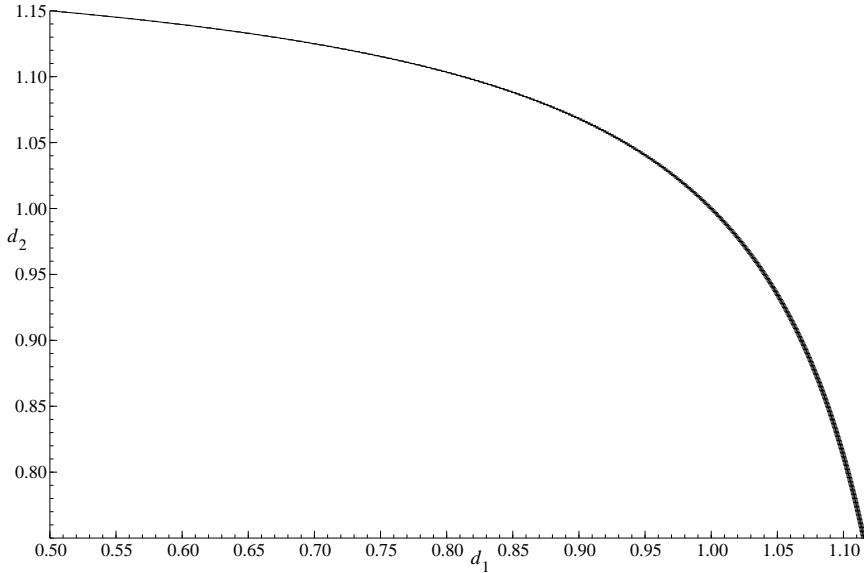

Figure 1: Stationarity region for two-state MS-GARCH with transition probabilities $\eta_{11} = \eta_{22} = 0.85$ and $d_i = \alpha_i + \beta_i$

As a special case, we consider a situation where we start the discrete Markov chain from its invariant distribution. This case is equivalent to a regime switching GARCH model where the probabilities are constant over time, see Bauwens, Preminger, and Rombouts (2006). Under assumptions A1-A2, it can be shown that a sufficient condition for geometric ergodictiy and existence of moments is given by $\sum_{j=1}^{n} \pi_j \mathrm{E}(\alpha_j u_t^2 + \beta_j)^k < 1$. We observe that the condition derived by Bollerslev (1986) for covariance-stationarity under a single GARCH model need not hold in each regime but for the weighted average of the GARCH parameters. Note, that high values of the parameters of the non-stable GARCH processes must be compensated by low probabilities for their regimes.

3 Estimation

Given the current computing capabilities, 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. 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.

3.1 Bayesian Inference

We explain the Bayesian algorithm for a MS-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.

For the case of two regimes, the model is given by equations (3)-(4), $s_t = 1$ or 2 indicating the active regime. We denote by Y_t the vector $(y_1\, y_2\, \ldots\, y_t)$ and likewise $S_t = (s_1\, s_2\, \ldots\, s_t)$. The model parameters consist of $\eta = (\eta_{11}, \eta_{21}, \eta_{12}, \eta_{22})'$, $\mu = (\mu_1, \mu_2)'$, and $\theta = (\theta_1', \theta_2')'$, where $\theta_k = (\omega_k, \alpha_k, \beta_k)'$ for $k = 1, 2$. The joint density of y_t and s_t given the past information and the parameters can be factorized as

$$f(y_t, s_t | \mu, \theta, \eta, Y_{t-1}, S_{t-1}) = f(y_t | s_t, \mu, \theta, Y_{t-1}, S_{t-1}) f(s_t | \eta, Y_{t-1}, S_{t-1}). \tag{5}$$

The conditional density of y_t is the Gaussian density

$$f(y_t | s_t, \mu, \theta, Y_{t-1}, S_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right) \tag{6}$$

where σ_t^2, defined by equation (4), is a function of θ. The marginal density (or probability mass function) of s_t is specified by

$$f(s_t | \eta, Y_{t-1}, S_{t-1}) = f(s_t | \eta, s_{t-1}) = \eta_{s_t s_{t-1}} \tag{7}$$

with $\eta_{11} + \eta_{21} = 1$, $\eta_{12} + \eta_{22} = 1$, $0 < \eta_{11} < 1$ and $0 < \eta_{22} < 1$. This specification says that s_t depends only on the last state and not on the previous ones and on the past observations of y_t, so that the state process is a first order Markov chain with no absorbing state.

The joint density of $y = (y_1, y_2, \ldots, y_T)$ and $S = (s_1, s_2, \ldots, s_T)$ given the parameters is then obtained by taking the product of the densities in (6) and (7) over all observations:

$$f(y, S | \mu, \theta, \eta) \propto \prod_{t=1}^{T} \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right) \eta_{s_t s_{t-1}}. \tag{8}$$

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 θ, μ, η, and the elements of S. We explain what are our prior densities for θ, μ, and η when we define the different blocks of the Gibbs sampler.

3.1.1 Sampling s_t

To sample s_t we must condition on s_{t-1} and s_{t+1} (because of the Markov chain for the states) and on the future state variables $(s_{t+1}, s_{t+2}, \ldots s_T)$ (because of path dependence of the conditional variances). The full conditional mass function of state t is

$$\varphi(s_t|S_{\neq t}, \mu, \theta, \eta, y) \propto \eta_{1,s_{t-1}}^{2-s_t} \eta_{2,s_{t-1}}^{s_t-1} \eta_{1,s_t}^{2-s_{t+1}} \eta_{2,s_t}^{s_{t+1}-1} \prod_{j=t}^{T} \sigma_j^{-1} \exp\left(-\frac{(y_j - \mu_{s_j})^2}{2\sigma_j^2}\right) \tag{9}$$

where we can replace $\eta_{2,s_{t-1}}$ by $1 - \eta_{1,s_{t-1}}$ and η_{2,s_t} by $1 - \eta_{1,s_t}$. Since s_t takes two values, it is easy to simulate this discrete distribution.

3.1.2 Sampling η

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

$$\varphi(\eta|S, \mu, \theta, y) \propto \pi(\eta) \prod_{t=1}^{T} \eta_{s_t s_{t-1}} \tag{10}$$

which does not depend on μ, θ and y. For simplicity, we can work with η_{11} and η_{22} as free parameters and assign to each of them a beta prior density on $(0,1)$. The posterior densities are then also independent beta densities. For example,

$$\varphi(\eta_{11}|S) \propto \eta_{11}^{a_{11}+n_{11}-1}(1 - \eta_{11})^{a_{21}+n_{21}-1} \tag{11}$$

where a_{11} and a_{21} are the parameters of the beta prior, n_{11} is the number of times that $s_t = s_{t-1} = 1$ and n_{21} is the number of times that $s_t = 2$ and $s_{t-1} = 1$. A uniform prior on $(0,1)$ corresponds to $a_{11} = a_{21} = 1$.

3.1.3 Sampling θ

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

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

which does not depend on η. We sample θ by 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 ω_1 given the values of S, β_1, α_1, θ_2, and μ sampled at iteration r, over a grid $(\omega_1^1, \omega_1^2 \cdots, \omega_1^G)$, to obtain the vector $G_\kappa = (\kappa_1, \kappa_2, \cdots, \kappa_G)$.

2. By a deterministic integration rule using M points, compute $G_f = (0, f_2, \ldots, f_G)$ where

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

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 θ_1 and θ_2 must be defined. The choice of these bounds (such as ω_1^1 and ω_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.

3.1.4 Sampling μ

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

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

which does not depend on η. It is not possible to factorize this function into the product of a function that depends on μ_1 but not on μ_2, and another one that depends on μ_2 but not on μ_1. Moreover, since σ_t depends on μ_1 or μ_2 (depending on t), the analytical form of the likelihood as a function of μ_1 or μ_2 is not a known density (e.g. a normal). Hence we must sample μ_1 and μ_2 jointly and numerically. We use the griddy-Gibbs sampler and we factorize the prior as a product of two uniform densities.

3.2 Simulation Example

We have simulated a data generating process (DGP) corresponding to the model defined by equations (3)-(4) for two states, and $u_t \sim N(0,1)$. The parameter values are reported in Table 2. The second GARCH equation implies a higher and more persistent conditional variance than the first one. The other parameter values are inspired by previous empirical results, like in Hamilton and Susmel (1994), and our results presented in the next section. In particular, the transition probabilities of staying in each regime are close to unity. All the assumptions for stationarity and existence of moments of high order are satisfied by this DGP. In Table 1, we report some summary statistics for 50,000 observations from this DGP, and in Figure 2, we show the 1,500 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 close to zero. The density is slightly skewed to the left, and its excess kurtosis is estimated to be 3.57 (the excess kurtosis is 1.62 for the first component GARCH and 0.12 for the second). The ACF of the squared data is strikingly more persistent than the ACF of each GARCH component, which are both virtually at 0 after 10 lags. 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 50000 simulated data

Mean	-0.001	Maximum	10.47
Standard deviation	1.623	Minimum	-12.76
Skewness	-0.156	Kurtosis	6.57

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 1,500 observations of the simulated data described above and shown in panel (a) of Figure 2. The results are in the last two columns of the table. In Figure 3, we report the corresponding posterior densities. The prior density of each parameter is uniform between the bounds reported in Table 2 with the DGP values. Thus, these bounds were used for the integrations in the griddy-Gibbs sampler (except for η_{11} and η_{22} since they are sampled from beta densities). The number of iterations of the Gibbs sampler was set to
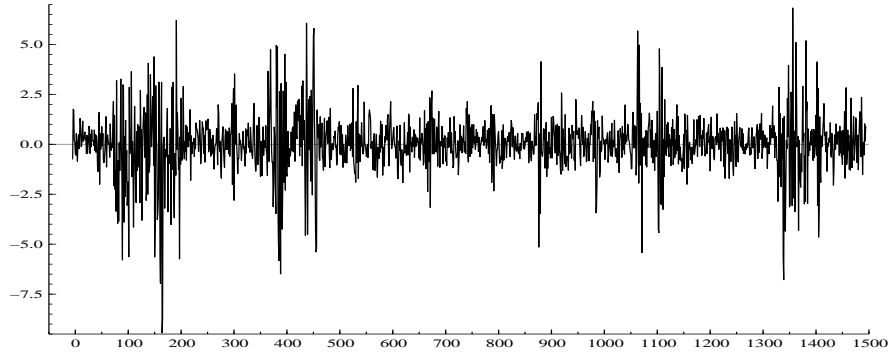
(a) Sample path (1500 observations)
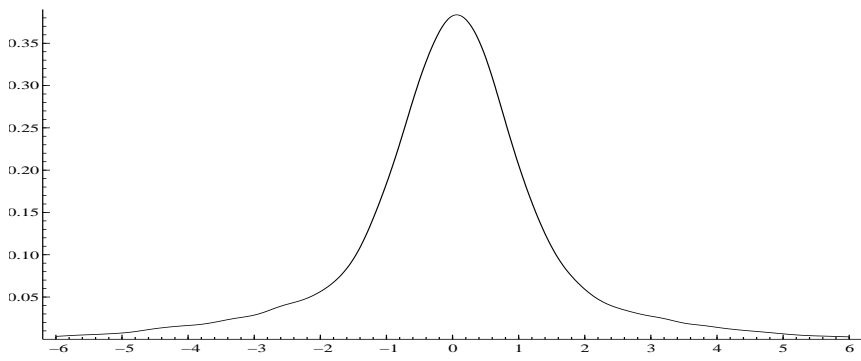
(b) Kernel density
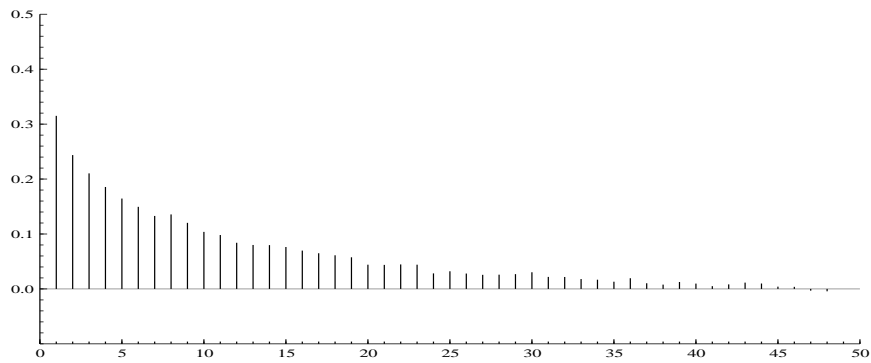
(c) Correlogram of squared data

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	Means	St. dev.
ω_1	0.30	(0.15 0.45)	0.345	(0.051)
β_1	0.20	(0.05 0.40)	0.192	(0.087)
α_1	0.35	(0.10 0.50)	0.264	(0.051)
ω_2	2.00	(0.50 4.00)	2.136	(0.688)
β_2	0.60	(0.35 0.85)	0.584	(0.106)
α_2	0.10	(0.02 0.35)	0.142	(0.049)
μ_1	0.06	(0.02 0.15)	0.079	(0.016)
μ_2	-0.09	(-0.35 0.18)	-0.076	(0.103)
η_{11}	0.98	(0.00 1.00)	0.987	(0.004)
η_{22}	0.96	(0.00 1.00)	0.958	(0.012)

Posterior means and standard deviations for MS-GARCH model. Sample of 1500 observations from DGP defined by equations (3)-(4) with $N(0,1)$ distribution.
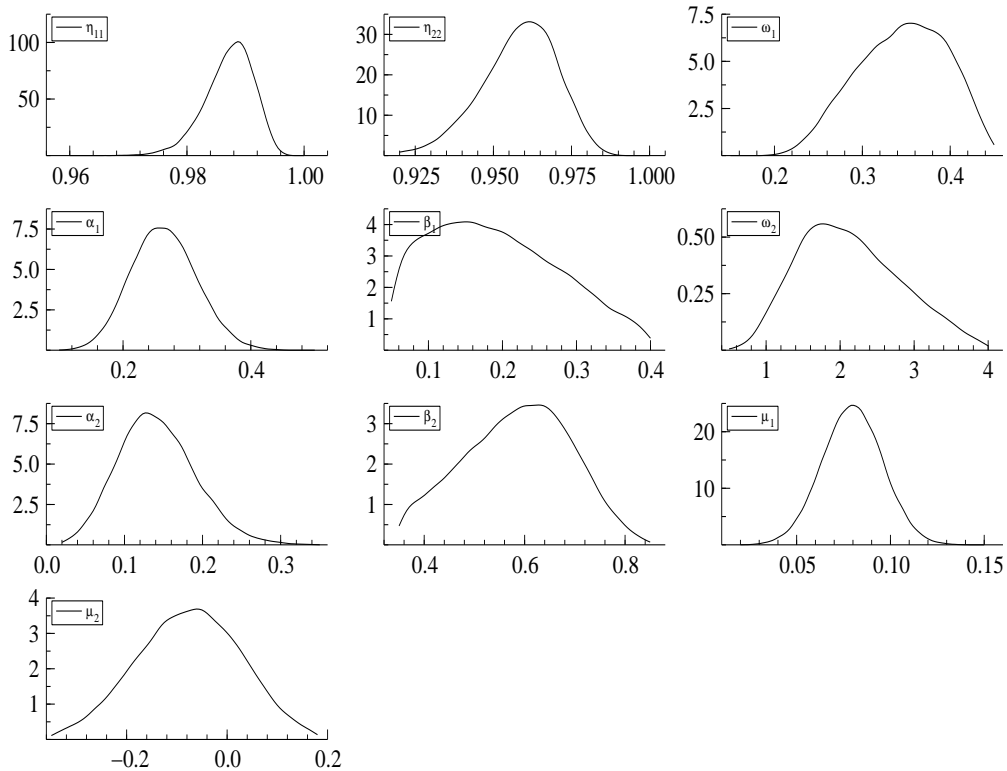

Figure 3: Posterior densities for the MS-GARCH model (simulation)

12

50,000, and the initial 20,000 draws were discarded, since after these the sampler seems to have converged (based on cumsum diagrams not shown to save space). Thus the posterior moments are based on 30,000 dependent draws of the posterior distribution. The posterior means are with few exceptions within less than one posterior standard deviation away from the DGP values, and the shapes of the posterior densities are not revealing bi-modalities that would indicate a label switching problem. From the Gibbs output, we also computed the posterior means of the state variables. These are obtained by averaging the Gibbs draws of the states. These means are smoothed (posterior) probabilities of the states. A mean state close to 1 corresponds to a high probability to be in the second regime. If we attribute an observation to regime 2 if its corresponding mean state is above one-half (and to regime 1 otherwise), we find that 96 per cent of the data are correctly classified.

4 Application

We use the S&P500 daily percentage returns from 19/07/2001 to 20/04/2007 (1500 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, a slightly positive skewness, and a usual excess kurtosis for this type of data and sample size, see also Table 3.

Table 3: Descriptive statistics for S&P500 daily returns

Mean	0.015	Minimum	-5.046
Standard deviation	1.00	Maximum	5.57
Skewness	0.11	Kurtosis	6.37

Sample period: 19/07/2001 to 20/04/2007 (1500 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 MS-ARCH model defined by setting $\beta_1 = \beta_2 = 0$ in equations (3)-(4), and a restricted version ($\beta_1 = \alpha_1 = 0$) of the corresponding MS-GARCH model. The marginal posterior densities for these models are shown in Figures 5 and 6. The intervals over which the densities are drawn are the prior intervals (except for the transition probabilities). The intervals for the
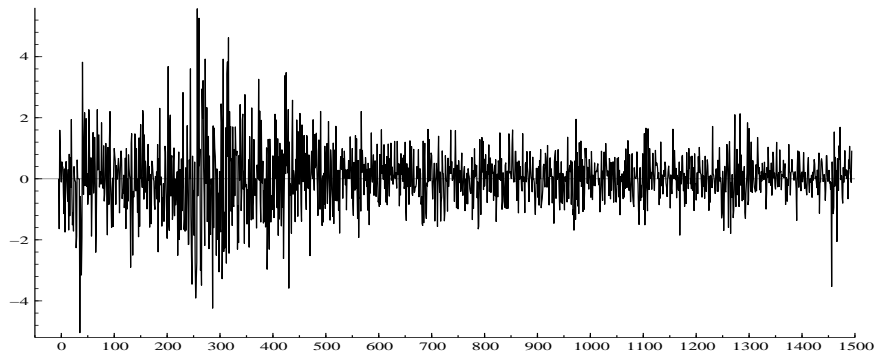
(a) Sample path (1500 observations)
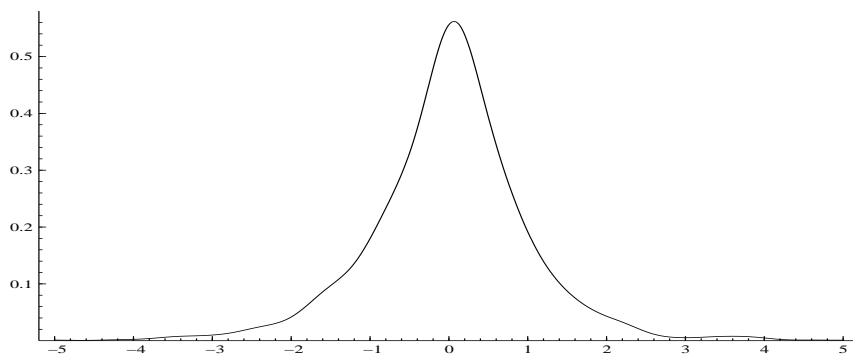
(b) Kernel density
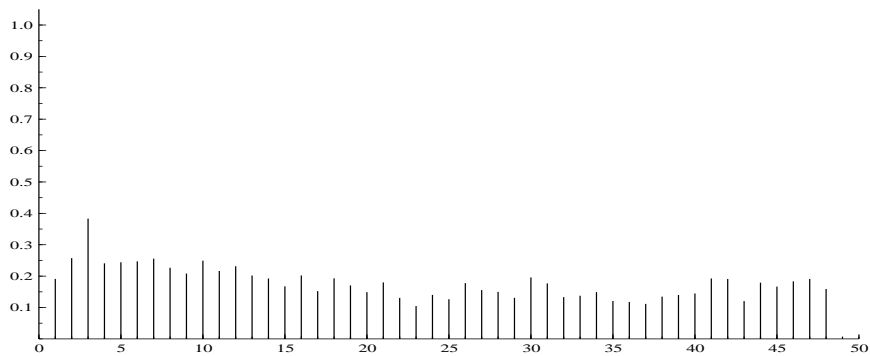
(c) Correlogram of squared data

Figure 4: Graphs for S&P500 daily returns from 19/07/2001 to 20/04/2007

Table 4: Posterior means and standard deviations (S&P500 daily returns)

	MS-ARCH		MS-GARCH	
	mean	st. dev.	mean	st. dev.
ω_1	0.419	(0.028)	0.308	(0.025)
β_1	-		-	
α_1	0.014	(0.012)	-	
ω_2	1.988	(0.175)	0.0467	(0.024)
β_2	-		0.919	(0.024)
α_2	0.115	(0.042)	0.054	(0.015)
μ_1	0.046	(0.016)	0.071	(0.019)
μ_2	-0.040	(0.044)	-0.012	(0.029)
η_{11}	0.994	(0.003)	0.978	(0.011)
η_{22}	0.986	(0.006)	0.985	(0.006)

Sample period: 19/07/2001 to 20/04/2007 (1500 observations). A - symbol means that the parameter was set to 0.

GARCH parameters were chosen to avoid negative values, and by trial and error so as to avoid truncation. The Gibbs sample size was fixed to 50,000 observations with a warm-up sample of 20,000, like for the simulation example.

When estimating the MS-ARCH model, we find that in the first regime, which is characterized by a low volatility level ($\omega_1/(1-\alpha_1) = 0.42$ using the posterior means as estimates, as opposed to 2.24 in the second regime), the ARCH coefficient α_1 is close to 0 (posterior mean 0.014, standard deviation 0.012, see also the marginal density in Figure 5). This is a weak evidence in favor of a dynamical effect in the low volatility regime. The same conclusion emerges after estimating the MS-GARCH model, with the added complication that the β_1 coefficient is poorly identified (since α_1 is almost null). Thus we opted to report the MS-GARCH results with α_1 and β_1 set equal to 0, and GARCH dynamics only in the high volatility regime. These results show clearly that the lagged conditional variance should be included in the second regime. Thus, the MS-ARCH model is not capturing enough the persistence of the conditional variance in the second regime. The second regime in the MS-GARCH model is rather strongly persistent but stable, with the posterior mean of $\beta_2 + \alpha_2$ equal to 0.973 ($0.919 + 0.054$). If we estimate a single regime GARCH model, we find that

the persistence is 0.990 $(0.942 + 0.048)$, which makes it closer to integrated GARCH than the second regime of the MS model. The estimation results for the MS-GARCH model also imply that compared to the first regime (where $\omega_1 = 0.31$), the second regime is a high volatility regime since $\omega_2/(1 - \alpha_2 - \beta_2) = 1.73$.
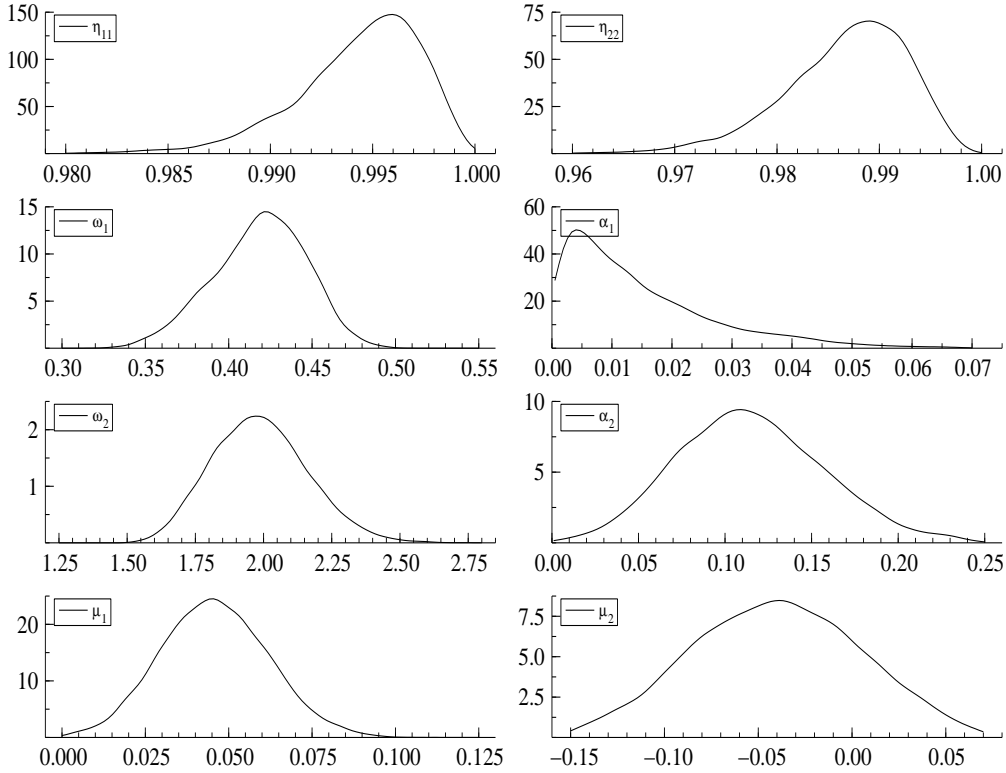

Figure 5: Posterior densities for the MS-ARCH model (S&P500 daily returns)

Another way to compare the two models is through the means of the state variables. These are obtained by averaging the Gibbs draws of the states. These means are smoothed (posterior) probabilities of the states. A mean state close to 1 corresponds to a high probability to be in the second regime. Figures 7 and 8 display the paths of these means. Both figures show, in conjunction with the sample path of the data (in Figure 4), that high probabilities are associated with high volatility periods (observations 1 to 500 and some peaks later). In this respect, the MS-ARCH model seems too insensitive in comparison with the GARCH version. From the posterior means of the MS-GARCH model, we can also deduce that the unconditional probabilities of the regimes are respectively 0.59 $(= (1 - \eta_{11})/(2 - \eta_{11} - \eta_{22}))$
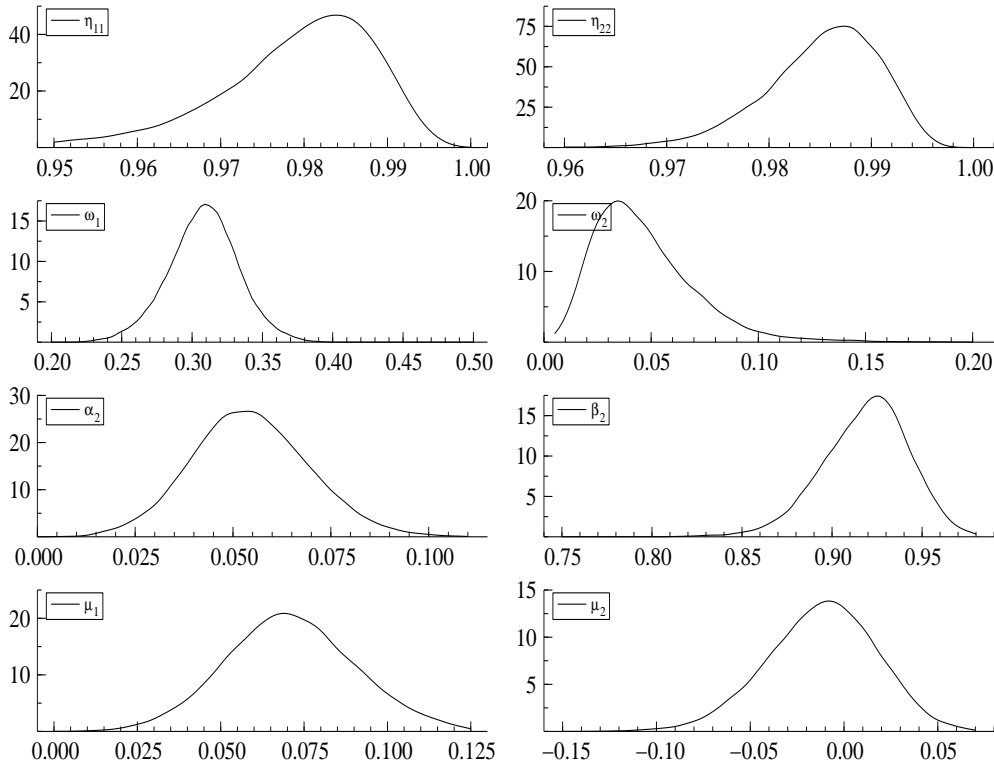
16

Figure 6: Posterior densities for the MS-GARCH model (S&P500 daily returns)

for the first one and 0.41 for the second one. These proportions correspond roughly to the information provided by the mean states.

5 Conclusion

We establish some theoretical properties for a Markov-switching univariate GARCH model with constant transition probabilities. We provide simple sufficient conditions for the ergodic stationarity of the process and the existence of its moments. We develop a reliable Bayesian estimation algortihm for this model, since ML estimation in not feasible due to path dependence.

Further research could be oriented in several directions. A first one consists in refining the specification by using existing extensions of the simple Gaussian GARCH model. A second direction of research is to specify the transition probabilities as a function of past information
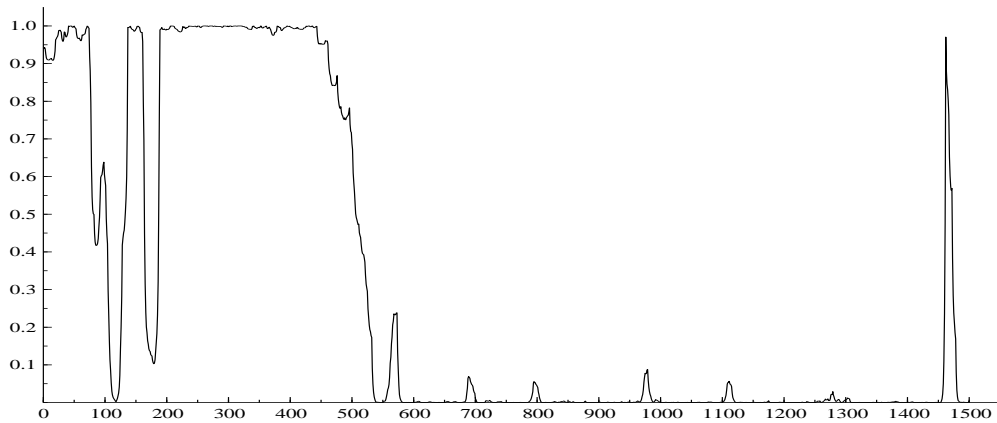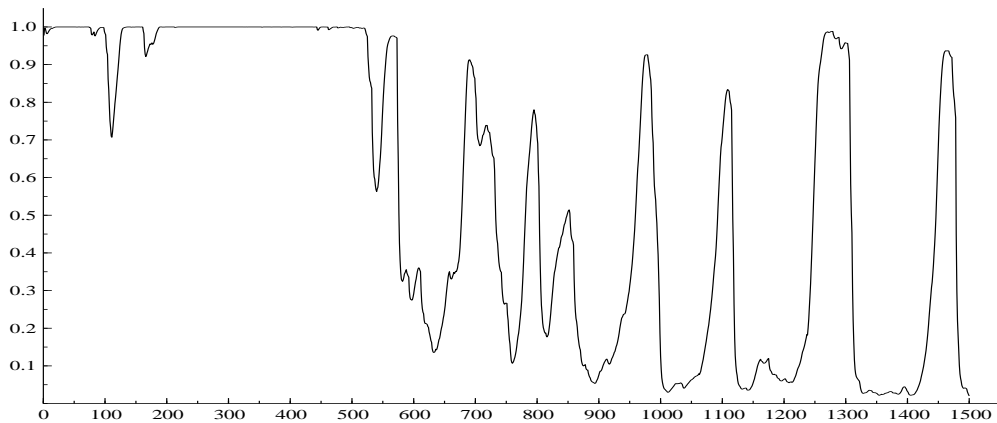
Figure 7: Mean states MS-ARCH

Figure 8: Mean states MS-GARCH

as in Gray (1996). These extensions would render the algorithm more CPU-time consuming (due to the additional parameters) but would not complicate it fundamentally. Establishing the ergodic stationarity and existence of moments of such more richly specified processes would require to extend and adapt the proofs presented in the current paper. Finally further research could be focussed on estimating the model with other data series, and on comparisons with other GARCH models, in a similar way as done by Marcucci (2005).

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 is a function of the information dated at time $t-1$ or earlier, not information dated at t. Let λ and v denote the Lebesgue and the counting measures, respectively.

Proof of Theorem 1: There exists a measurable function $g : S \times \Re \rightarrow S$ such that $s_t = g(s_{t-1}, \xi_t)$, where the error term ξ_t is i.i.d. independent of u_t and h_0. Let $\eta_t = (u_t, \xi_t)'$ and Z_t is defined on $D \subset \Re \times \Re_+ \times \Re$ where $\Re_+ = (0, +\infty)$. From (3) and (4), we have

$$
Z_t = \begin{pmatrix} y_t \\ h_t \\ s_t \end{pmatrix} = \begin{pmatrix} \mu_{g(s_{t-1},\xi_t)} + \sqrt{h_{t-1}} u_t \\ \omega_{g(s_{t-1},\xi_t)} + \alpha_{g(s_{t-1},\xi_t)} \varepsilon_t^2 + \beta_{g(s_{t-1},\xi_t)} h_{t-1} \\ g(s_{t-1}, \xi_t) \end{pmatrix} \tag{15}
$$

$$
= \begin{pmatrix} \mu_{g(s_{t-1},\xi_t)} + \sqrt{h_{t-1}} u_t \\ \omega_{g(s_{t-1},\xi_t)} + (\alpha_{g(s_{t-1},\xi_t)} u_t^2 + \beta_{g(s_{t-1},\xi_t)}) h_{t-1} \\ g(s_{t-1}, \xi_t) \end{pmatrix} = F(Z_{t-1}, \eta_t)
$$

where $F : D \times \Re^2 \rightarrow D$. Since η_t is independent of Z_{t-1} its follows from (15) that $(y_t, h_t, s_t)'$ forms a homogeneous Markov chain.

The process is defined on $(D, \Im, \varphi)$. The state space of the process is given by $D = \{(y, h, s) \in \Re \times \Re_+ \times \Re : (y, h) \in \bigcup_{i=1}^n D_i, \; s \in S\}$, where D_i is the domain of the chain in each regime and is given by $D_i = \{(y, h) \in \Re \times \Re_+ : h \geq \omega_i + \alpha_i(y - \mu_i)^2 + \beta_i \bar{h}\}$ and $\bar{h} = \min_{\beta_i < 1}\{\omega_i/(1 - \beta_i)\}$ (see Zhang, Russell, and Tsay (2001)). The strict stationarity of the first regime $(E\log(\alpha_1 u_t^2 + \beta_1) < 0)$, implies that $\beta_1 < 1$ (see, Nelson (1990)), hence $\bar{h}$ is well defined. The state space is equipped with $\Im$, the Borel σ−algebra on $\Re \times \Re_+ \times \Re$ restricted to D. The measure φ is the product measure $\lambda^2 \otimes v$ on $(D, \Im)$. We use $P^m(z_0, A) = P(Z_t \in A | Z_{t-m} = z_0)$ to signify the probability that (y_t, h_t, s_t) moves from (y_0, h_0, s_0) to the set $A \in \Im$ in m steps.

In order to establish the geometric ergodicity of the Markov chain, we first show that the process is φ−irreducible. For irreducibility, it is sufficient to show that $P^k(z_0, A) > 0$ for

some $k \geq 1$, for all $z_0 \in D$ and any Borel measurable sets $A \in \Im$ with positive φ measure (see Chan (1993)). We can show that from any $(y_0, h_0, s_0) \in D$, all $(y, h, s) \in A$ can be reached in a finite number of steps. We assume that $s_0 = i$, $s = \ell$ and $\bar{h}$ is achieved in regime q. Let $\tilde{h} = [h - \omega_\ell - \alpha_\ell (y - \mu_\ell)^2]/\beta_\ell$ and $\varsigma = \tilde{h} - \bar{h}$; since $\tilde{h} > \bar{h}$ we have $\varsigma > 0$. Thus, there exists a positive integer $m = \min(t \geq 1 : \bar{h} + 0.5\varsigma + 0.5\beta_q^{-t}\varsigma > \omega_1 + \beta_1 h_0\}$, such that the point (y, h, s) can be reached through the following $m + 1$ intermediate steps: $w = \{(y_t, h_t, s_t)\}_{t=1}^{m+1}$ where $s_1 = 1$, $s_t = q$, $h_t = \bar{h} + 0.5\varsigma + 0.5\beta_q^{i-(m+1)}\varsigma$, $y_1 = \mu_1 + [(h_1 - \omega_1 - \beta_1 h_0)/\alpha_1]^{0.5}$, $y_t = \mu_q + [0.5\varsigma(1 - \beta_q)/\alpha_q]^{0.5}$ for $t \leq m + 1$ and in the $m + 2$-th step $(y_{m+1}, \tilde{h}, q) \to (y, h, \ell)$. The $m+2$-th step transition probability is absolutely continuous with respect to the φ measure. Thus, $\mathrm{P}^{m+2}(z_0, A) = \int_A p^{m+2}(z_0, z)d\varphi(z)$, and by assumptions A1 and A2,

$$p^{m+2}(z_0, z) \geq \prod_{i=0}^{m+1} f((y_{i+1} - \mu_{s_{i+1}})/h_i^{0.5})\mathrm{P}(s_{i+1}|s_i) > 0,$$

which implies that $\mathrm{P}^{m+2}(z_0, A) > 0$ and hence the chain is $\varphi-$irreducible. If $z_0 \in C$, a compact set, $\inf_{(z_0,z)\in C\times C} p^{m+2}(z_0, z) > \delta > 0$ and for any $A \in \Im$ and $z_0 \in C$,

$$\mathrm{P}^{m+2}(z_0, A) \geq \mathrm{P}^{m+2}(z_0, A \cap C) \geq \int_{A\cap C} p^{m+2}(z_0, z)d\varphi(z) \geq \delta\varphi(A \cap C).$$

Therefore, $\mathrm{P}(z_0, A)$ is minorized by $\varphi(\cdot \cap C)$ which implies that all non-null, compact sets in D are small by definition, see Meyn and Tweedie (1993, p. 111), and can serve as test sets. Using the same arguments as above, we can show that any small set can be reached in $m + 3$ steps, therefore the chain is aperiodic, see Chan (1993).

From (4) and the c_r inequality we get

$$h_t^{1/t} \leq [\omega_{s_t} + (\alpha_{s_t} u_t^2 + \beta_{s_t})h_{t-1}]^{1/t} \tag{16}$$

$$\leq (\omega_{s_t})^{1/t} + (\alpha_{s_t} u_t^2 + \beta_{s_t})^{1/t}(\omega_{s_{t-1}})^{1/t} + [(\alpha_{s_t} u_t^2 + \beta_{s_t})(\alpha_{s_{t-1}} u_{t-1}^2 + \beta_{s_{t-1}})]^{1/t}h_{t-2}^{1/t}$$

$$\vdots$$

$$\leq \prod_{j=1}^t (\alpha_{s_j} u_j^2 + \beta_{s_j})^{1/t}h + \sum_{j=0}^{t-1} (\omega_{s_{t-j-1}})^{1/t}\prod_{i=0}^j (\alpha_{s_{t-i}} u_t^2 + \beta_{s_{t-i}})^{1/t} + (\omega_{s_t})^{1/t}$$

Since $\{s_t\}$ is an ergodic Markov chain, for any initial state, we have that

$$\frac{1}{t}\sum_{j=1}^t \log(\alpha_{s_j} u_j^2 + \beta_{s_j}) \to \sum_{i=1}^n \pi_i \mathrm{E}[\log(\alpha_i u_t^2 + \beta_i)] < 0, \text{a.s.} \tag{17}$$

see Chan (1993). This result, with assumption A1 and the dominated convergence theorem imply that there exists a $\bar{t}$ sufficiently large such that $\delta \geq 1/\bar{t} = p$ and for all $j \in S$,

$$E\left(\prod_{j=1}^{t}(\alpha_{s_j}u_j^2 + \beta_{s_j})^p|s_0 = j\right) = \gamma < 1. \tag{18}$$

As a drift function we use $V(z) = 1 + (\bar{\eta}/\Delta)y^{2p} + h^p$, where $\bar{\eta} = \eta - \gamma$, $\Delta = E(u_t^{2p})$, $p = 1/\bar{t}$ and η is some positive number which satisfies $\gamma < \eta < 1$ and the test set is given by $C = \{(y, h, s) \in D : h + y^2 \leq c, \ s \in S\}$, where $c > 0$ is to be determined below.

From (16)-(18), we find

$$E(h_t^p|z_0 = z) \leq 1 + E\left(\prod_{j=1}^{t}(\alpha_{s_j}u_j^2 + \beta_{s_j})^p|s_0 = j\right)h^p + M \leq M + \gamma h^p$$

where $M = 1 + E\left(\sum_{j=0}^{t-1}(\omega_{s_{t-j-1}})^p\prod_{i=0}^{j}(\alpha_{s_{t-i}}u_t^2 + \beta_{s_{t-i}})^p + (\omega_{s_t})^p|s_0 = j\right)$ and $M < \infty$ by assumption A1. Therefore,

$$E(V(z_t)|z_0 = z) \leq 1 + M + \gamma h^p + \bar{\eta}E(h_t^p|z_0 = z)E|u_t^{2p}|)/\Delta \leq h^p\left(\frac{M}{h^p} + \eta\right)$$

Since the Lyapounov function above is bounded on compact sets and $h < V(z)$, we can choose c and $\eta' \in (\eta, 1)$ such that $E(V(z_t)|z_0 = z) \leq \eta' \cdot V(z) + a \cdot 1_C(z)$ for some $a < \infty$ and for all $z \in D$, hence the drift criterion is satisfied. We can then combine Meyn and Tweedie (1993, Theorem 15.0.1) and Tjostheim (1990) to obtain that $\{Z_t\}$ is geometrically ergodic and so is $\{y_t\}$. The finiteness of $E(|Z_t|^p)$ with respect to the stationary measure follows from Meitz and Saikkonen (2006, Lemma 6). If the process is initiated from its stationary distribution, it further follows that the process is β-mixing with geometrically decaying mixing numbers, see Doukhan (1994, p.89).

Proof of Theorem 2: Let I_j be an $1 \times n$ matrix that contain 1 on the j-th position and zeros elsewhere and $l = (1, \ldots, 1)$ be an $n \times 1$ vector. The matrix Ω is positive, hence the spectral radius is real and positive and assumption A5 implies that there exists a positive integer r such that each element of Ω^r is smaller than $1/n$, that is $(\Omega^r)_{ij} < 1/n$, see Lutkepohl (1996, p. 76, 3(a)), hence

$$E\left(\prod_{j=1}^{r}(\alpha_{s_j}u_j^2 + \beta_{s_j})h \,|Z_0 = (y, h, j)\right) = (I_j\Omega)\Omega^{r-1}lh = \gamma < 1.$$

21

By solving (4) recursively and setting $h_0 = h$, we get

$$h_t = [\omega_{s_t} + (\alpha_{s_t} u_t^2 + \beta_{s_t}) h_{t-1}] = \prod_{j=1}^{r} (\alpha_{s_j} u_j^2 + \beta_{s_j}) h + \sum_{j=0}^{r-1} \omega_{s_{t-j-1}} \prod_{i=0}^{j} (\alpha_{s_{t-i}} u_t^2 + \beta_{s_{t-i}}) + \omega_{s_t}.$$
(19)

Let $\Delta = E(u_t^{2k})$ and $\bar{\eta} = \eta - \gamma^k$ where η is some positive number which satisfies $\gamma^k < \eta < 1$. We select a drift function of the form $V(y, h, s) = 1 + (\bar{\eta}/\Delta) y^{2k} + h^k$ and a test set $C = \{(y, h, s) \in D : h + y^2 \le c, \ s \in S\}$. By the binomial theorem, (19), assumption A4 and some tedious calculations, we find

$$E(V(y_t, h_t) | Z_0 = (y, h, j)) = 1 + E(h_t | z) + \bar{\eta} E(y_t^{2k} | z)/\Delta =$$

$$1 + E\left(\prod_{j=1}^{r} (\alpha_{s_j} u_j^2 + \beta_{s_j})^j\right)^k h^k + \bar{\eta} h^k + o(h^k) \le h^k(\gamma^k + \bar{\eta} + o(1)) \le h^k(\eta + o(1))$$

By applying the same arguments as in Theorem 1, the desired result follows.

References

ABRAMSON, A., AND I. COHEN (2007): "On the stationarity of Markov-switching GARCH processes," *Econometric Theory*, 23, 485–500.

BAUWENS, L., A. PREMINGER, AND J. ROMBOUTS (2006): "Regime switching GARCH models," CORE DP 2000/11, Université catholique de Louvain, Louvain La Neuve.

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., R. ENGLE, AND D. NELSON (1994): "ARCH Models," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, chap. 4, pp. 2959–3038. North Holland Press, Amsterdam.

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.

DAS, D., AND B. H. YOO (2004): "A Bayesian MCMC algorithm for Markov switching GARCH models," Manuscript, City UNiversity of New York and Rutgers University.

DIEBOLD, F. (1986): "Comment on Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 51–56.

DOUKHAN, P. (1993): *Mixing: Properties and Examples.* Springer-Verlag, New-York.

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.-M. ZAKOIAN (2001): "Conditional heteroskedasticity driven by hidden Markov chains," *Journal of Time Series Analysis*, 22, 197–220.

FRANCQ, C., AND J.-M. ZAKOIAN (2002): "Comments on the paper by Minxian Yang: "Some properties of vector autoregressive processes with Markov-switching coefficients"," *Econometric Theory*, 18, 815–818.

——— (2005): "The L^2-structures of standard and switching-regime GARCH models," *Stochastic Processes and their Applications*, 1158, 1557–1582.

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 (2004): "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.

HENNEKE, J. S., S. T. RACHEV, AND F. J. FABOZZI (2006): "MCMC based estimation of Markov Switching ARMA-GARCH models," Working Paper, University of Karlsruhe.

KAUFMAN, S., AND S. FRUHWIRTH-SCHNATTER (2002): "Bayesian analysis of switching ARCH models," *Journal of Time Series Analysis*, 23, 425–458.

KAUFMAN, S., AND M. SCHEICHER (2006): "A switching ARCH model for the German DAX index," *Studies in Nonlinear Dynamics and Econometrics*, 10/4, Article 3.

KLAASSEN, F. (2002): "Improving GARCH Volatility Forecasts with Regime-Switching GARCH," *Empirical Economics*, 27, 363–394.

LAMOUREUX, C., AND W. LASTRAPES (1990): "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects," *Journal of Finance*, 45, 221–229.

LUTKEPOHL, H. (1996): *Handbook of Matrices*. Wiley, New-York.

MARCUCCI, J. (2005): "Forecasting stock market volatility with regime-switching GARCH models," *Studies in Nonlinear Dynamics and Econometrics*, 9, 1–53.

MEITZ, M., AND P. SAIKKONEN (2006): "Stability of nonlinear AR-GARCH models," SSE/SFI Working paper series in Economics and Finance No. 632.

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.

NELSON, D. B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

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.

TJOSTHEIM, D. (1990): "Non-linear time series and Markov chains," *Advances in Applied Probability*, 22, 587–611.

YANG, M. (2000): "Some properties of vector autoregressive processes with Markov-switching coefficients," *Econometric Theory*, 16, 23–43.

YAO, J. (2001): "On square-integrability of an AR process with Markov switching," *Statistics and Probability Letters*, 52, 265–270.

YAO, J., AND J.-G. ATTALI (2000): "On stability of nonlinear AR process with Markov switching," *Advances in Applied Probability*, 32, 394–407.

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.