

Theory and inference for a Markov switching GARCH model

LUC BAUWENS[†], ARIE PREMINGER[‡] AND JEROEN V. K. ROMBOUTS^{†,§,¶}

[†]*Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve, Belgium.*

E-mail: luc.bauwens@uclouvain.be

[‡]*Department of Economics, University of Haifa, 31905, Israel.*

E-mail: ariepr@bgu.ac.il

[§]*HEC Montréal and CIRPEE, 3000 Cote Sainte Catherine,
Montréal (Quebec), Canada, H3T 2A7.*

E-mail: jeroen.rombouts@hec.ca

[¶]*CIRANO, 2020, University Street, 25th Floor, Montréal, Quebec, Canada, H3A 2A5.*

First version received: June 2008; final version accepted: November 2009

Summary 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 S&P500 daily returns.

Keywords: Bayesian inference, GARCH, Markov-switching.

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 et al., 1994, and Giraitis et al., 2007, for a review of more recent developments). 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 behaviour 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) and Hilebrand (2005) 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) considers 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. Cai (1994) and Hamilton and Susmel (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. In his model, the path dependence problem is removed by aggregating the conditional variances over all regimes at each time step in order to construct a single variance term. This term (conditional on the available information, but not on the regimes) is used to compute the conditional variances in the next time step. A modification of his model is suggested by Klaassen (2002); see also Dueker (1997), Bollen et al. (2000), Haas et al. (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-switching 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. Hence, in our model the conditional variance at each time depends on the whole regime path. This constitutes the main difference between our model and existing variants of Gray's (1996) model mentioned above. 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. A similar model was proposed by Francq and Zakoian (2005) who study conditions for second-order stationarity and existence of higher-order moments.

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 repeated 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 et al. (2001) establish the consistency of the ML estimator. Douc et al. (2004) obtained its asymptotic normality for a class of autoregressive models with Markov regime, which includes the regime-switching ARCH model as a special case. 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 Frühwirth-Schnatter (2002) and Kaufman and Scheicher (2006). Tsay (2005, pp. 588–94) proposed a Bayesian approach for a simple two-state Markov switching model with different risk premiums and different GARCH dynamics. Das and Yoo (2004) and Gerlach and Tuyl (2006) 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 only the last cited paper contains an application to real data. Finally, the most comparable work to our paper (for estimation) is that of Henneke et al. (2006) who

estimate by an 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. To illustrate our estimation method, we apply it to a time series of daily returns of the S&P500 index. We find that the MS-ARCH version of the model does not account satisfactorily for the persistence of the conditional variance, whereas the MS-GARCH version performs better. Moreover, the latter dominates the former in terms of a model choice criterion based on the BIC formula.

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. 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

$$y_t = \mu_t + \sigma_t u_t \quad (2.1)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2.2)$$

where σ_t and μ_t are measurable functions of $y_{t-\tau}$ for $\tau \leq t-1$, $\varepsilon_t = y_t - \mu_t$, 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 (2.2). 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 Markov-switching GARCH (MS-GARCH) model that permits regime switching in the parameters. 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 a special case. Let $\{s_t\}$ be an ergodic homogeneous Markov chain on a finite set $S = \{1, \dots, n\}$, with transition matrix P defined by the probabilities $\{\eta_{ij} = P(s_t = i | s_{t-1} = j)\}$ and invariant probability measure $\pi = \{\pi_i\}$. We assume the chain is initiated at $t = 0$, which implies that $\{s_t\}_{t \geq 1}$ are independent by definition from $\{u_t\}_{t \geq 1}$ since the transition probabilities are fixed over time. The MS-GARCH model is defined by

$$y_t = \mu_{s_t} + \sigma_t u_t \quad (2.3)$$

$$\sigma_t^2 = \omega_{s_t} + \alpha_{s_t} \varepsilon_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2, \quad (2.4)$$

where $\omega_{s_t} > 0$, $\alpha_{s_t} \geq 0$, $\beta_{s_t} \geq 0$ for $s_t \in \{1, 2, \dots, n\}$, and $\varepsilon_t = y_t - \mu_{s_t}$. These assumptions on the GARCH coefficients entail that σ_t^2 is almost surely strictly positive. 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. However, these results do not apply in our framework since, although for the GARCH model the squared residuals follow an ARMA model, the same does not apply for our model. 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 et al. (2001). The MS-GARCH process is not a Markov chain. However, the extended process $Z_t = (y_t, \sigma_{t+1}^2, s_{t+1})'$ is a Markov chain (see the Appendix). In what follows, we state mild regularity conditions for which this chain is geometrically ergodic, strictly stationary and has finite moments. These issues are not dealt with by Francq and Zakoian (2005). Further, geometric ergodicity implies that it is not necessary for our model to have a stationary solution, but also that irrespective of the initial conditions it asymptotically behaves as the stationary version. In this case, the asymptotic analysis can be extended to any solution of the model, and not only the stationary one. Our results are based on Markov chain theory; see e.g. Chan (1993) and Meyn and Tweedie (1993). We impose the following assumptions:

ASSUMPTION 2.1. *The error term u_t is i.i.d. with a density function $f(\cdot)$ that is positive and continuous everywhere on the real line and is centred on zero. Furthermore, $E(|u_t^2|^\delta) < \infty$ for some $\delta > 0$.*

ASSUMPTION 2.2. *$\alpha_i > 0$, $\beta_i > 0$, the Markov chain is homogeneous, and $\eta_{ij} \in (0, 1)$ for all $i, j \in \{1, \dots, n\}$.*

ASSUMPTION 2.3. $\sum_{i=1}^n \pi_i E[\log(\alpha_i u_t^2 + \beta_i)] < 0$.

The first assumption is satisfied for a wide range of distributions for the error term, e.g. the normal and the Student distributions. For $\delta \geq 1$, 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 2.3 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 $E \log(\alpha_1 u_t^2 + \beta_1) < 0$. To obtain the results in Theorem 2.1, 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 2.1. *Under Assumptions 2.1–2.3, Z_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, $E(|y_t|^{2p}) < \infty$ for some $p \in (0, \delta]$ where the expectations are taken under the stationary distribution.*

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. Geometric ergodicity implies that if the process is initiated from its stationary distribution, it is regular mixing—see Doukhan (1994, sec. 1.1) for the definition—with exponential decaying mixing numbers. This implies that the autocovariance function of any measurable function of the data (if it exists) converges to zero at least at the same rate (e.g. the autocorrelation function of $|y_t|^p$ decays at exponential rate). Using geometric ergodicity, we can evaluate the stationary distribution numerically since analytical results are hard to obtain for our model. For example, Moeanaddin and Tong (1990) propose a conditional density approach by exploiting the Chapman–Kolmogorov relation for non-linear models; see also Stachurski and Martin (2008).

For the GARCH(1,1) model, the sum $\alpha + \beta$ measures the persistence of a shock to the conditional variance. We note that (2.4) can be rewritten as

$$\sigma_t^2 = \omega_{s_t} + \lambda_t \sigma_{t-1}^2 + v_t, \quad (2.5)$$

where $\lambda_t = \alpha_{s_t} + \beta_{s_t}$ and $v_t = \alpha_{s_t}(\varepsilon_{t-1}^2 - \sigma_{t-1}^2)$ is a serially uncorrelated innovation. Let $A_k = \prod_{j=1}^k \lambda_{t-j}$ for $k \geq 1$ ($A_0 \equiv 1$), $\gamma = \sum_{i=1}^n \pi_i \log(\alpha_i + \beta_i)$. Assuming $\gamma < 0$, by solving (2.5) recursively, we can express the conditional variance as

$$\sigma_t^2 = \sum_{k=0}^{t-1} A_k(v_{t-k} + \omega_{s_{t-k}}) + A_t \sigma_0^2.$$

It can be shown that $A_k = O(\exp(\bar{\gamma}k))$ for some $\bar{\gamma} \in (\gamma, 0)$ with probability one.¹ Thus, the impact of past shocks on the variances declines geometrically at a rate which is bounded by γ . Hence, γ can serve as a bound on the measure of volatility persistence of our model. As it approaches zero the persistence of the volatility shocks increases. For $\gamma = 0$, the impact of shocks on the variances does not decay over time. By Jensen's inequality and the strict concavity of $\log(x)$, it is clear that if $\gamma \leq 0$ Condition 2.3 is satisfied.

Next, we illustrate that Condition 2.3 allows explosive regimes while the global process is stable. We consider an MS-GARCH model with two regimes. For the first regime we choose an integrated GARCH(1,1) process with realistic values $\alpha_1 = 0.1$ and $\beta_1 = 0.9$. We note that this process is strictly stationary, but not covariance stationary. Let $\pi = \eta_{21}/(\eta_{12} + \eta_{21})$ be the ergodic probability of the stable regime, and let (α_2, β_2) be the parameters of the second regime. Further, define $F(\alpha_2, \beta_2, \pi) = \pi E \log(0.1u_t^2 + 0.9) + (1 - \pi) E \log(\alpha_2 u_t^2 + \beta_2)$. In Figure 1, we show the strict stationarity frontiers $F(\alpha_2, \beta_2, \pi) = 0$ which have been evaluated by simulations in the cases $u_t \sim N(0, 1)$ and $\pi = 0.0, 0.5, 0.75$. The strict stationarity regions are the areas below these curves and above the axes (notice that on the graph $\beta_2 \geq 0.80$). We note that when $\pi = 0$ the model has one regime which is strictly stationary and the computed values satisfy the stability condition of Nelson (1990). However, for $\pi > 0$, the parameters of the second regime imply that it can be explosive. That is, under the non-stable regime the conditional volatility diverges. Further, the higher the probability of being in the stable regime, the higher the values that the persistence parameters of the second regime can assume. Therefore, we observe that our model allows periods in which explosive regimes are operating, giving the impression of structural instability in the conditional volatility, before the process collapses to its stationary level.

¹ Note that $\{s_t\}$ is an ergodic Markov chain, hence for any initial state, $\frac{1}{k} \sum_{j=1}^k \log(\alpha_{s_j} + \beta_{s_j}) \rightarrow \gamma$ a.s.; hence similar to Nelson (1990, Theorem 2) we can show that there exists a $\bar{\gamma} \in (\gamma, 0)$ such that $A_k = O(\exp(\bar{\gamma}k))$ with probability one.

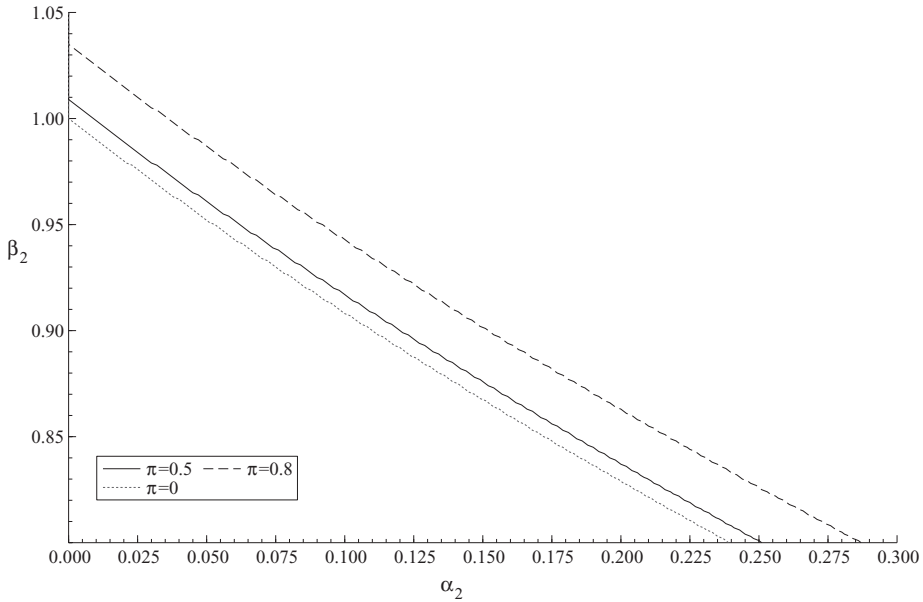


Figure 1. Strict stationarity region for a two-state MS-GARCH process.

In order to establish **the existence of higher-order moments**, we define the $n \times n$ matrix

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

A similar matrix was first introduced by Yao and Attali (2000) for non-linear 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:

ASSUMPTION 2.4. $E(|u_t^2|^k) < \infty$ for some integer $k \geq 1$.

ASSUMPTION 2.5. $\rho(\Omega) < 1$.

Assumption 2.5 is similar to the stability condition imposed by Francq and Zakoian (2005) to establish the existence of a second-order stationary solution for the MS-GARCH model. However, in our set-up this condition induces not only stationarity but also geometric ergodicity.

THEOREM 2.2. *Under Assumptions 2.1–2.2 and 2.4–2.5, 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 2.2 is simple to check in the leading case where $k = 1$. Let $d_i = \alpha_i + \beta_i$. If $d_i < 1$ for all $i \in \{1, 2, \dots, n\}$, Assumption 2.5 is satisfied for this case, since $\eta_{ij} \in (0, 1)$, see Lutkepohl (1996, p. 141, 4(b)), and the resulting MS-GARCH

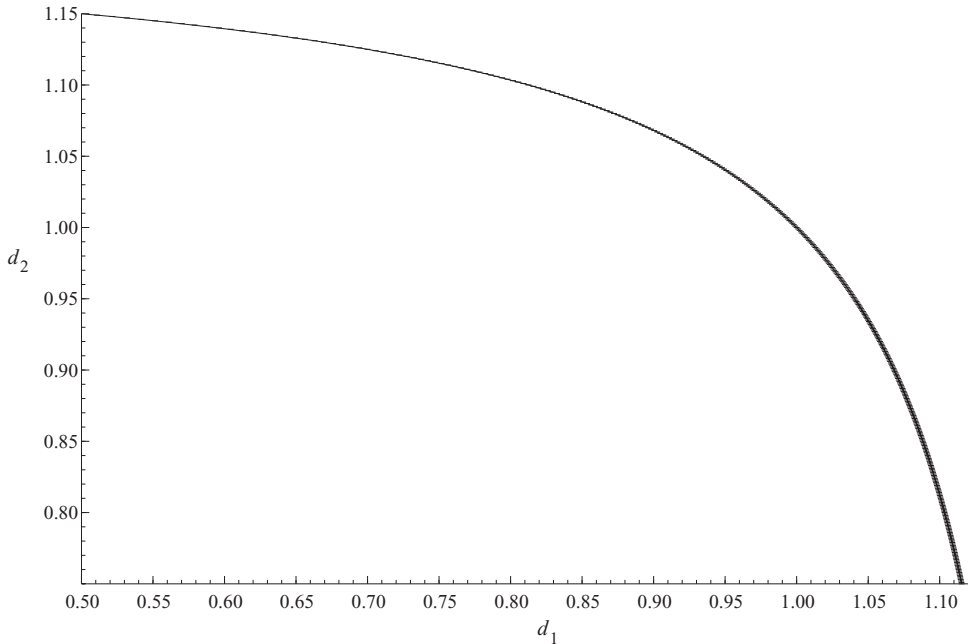


Figure 2. Covariance-stationarity region for two-state MS-GARCH.

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 2 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.

As a special case, we consider a situation where we start the discrete ergodic Markov chain from its invariant measure π . In this case $\pi = P\pi$ and our model is equivalent to a regime switching GARCH model where the probabilities of the regimes are constant over time; see Bauwens et al. (2006). Under Assumptions 2.1–2.2, it can be shown that a sufficient condition for geometric ergodicity and existence of moments of order k is given by $\sum_{i=1}^n \pi_i E(\alpha_i u_i^2 + \beta_i)^k < 1$. We observe that the condition derived by Bollerslev (1986) for covariance stationarity under a single GARCH model needs 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.

The autocorrelation function of ε_t^2 exists if Assumptions 2.4 and 2.5 are satisfied for $k = 2$, and the geometric ergodicity implies that it decreases exponentially. So, our model allows for short memory in the squared residuals. Now, let ψ be the second largest eigenvalue of P , which is assumed to be homogeneous in our setting. Kramer (2008) shows that if $\rho(\Omega) < 1$ (2.5) and P is non-homogeneous and depends on the sample size (T) such that $1 - \psi = O(T^{-2d})$ for some $d > 0$ then $\text{Var}(\sum_{t=1}^T \varepsilon_t^2) = O(T^{2d+1})$. This implies that the process has a long memory in the squared residuals; see also Diebold and Inoue (2001).

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. It is applied for Bayesian inference on stochastic volatility models, as initially proposed by Jacquier et al. (1994); see also Bauwens and Rombouts (2004) for a survey including other references. In Section 3.1, we present the Gibbs sampling algorithm for the case of two regimes, in Section 3.2 we discuss the relation between identification and the prior, in Section 3.3 we discuss possible extensions of the algorithm and related issues, and in Section 3.4 we perform a Monte Carlo study using a realistic data-generating process.

3.1. Bayesian inference

We explain the Gibbs sampling algorithm for an 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 sceptical 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 (2.3)–(2.4), $s_t = 1$ or 2 indicating the active regime. We denote by Y_t the vector $(y_1 y_2 \dots y_t)$ and likewise $S_t = (s_1 s_2 \dots s_t)$. The 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 factorised as

$$f(y_t, s_t | \mu, \theta, \eta, Y_{t-1}, S_{t-1}) = f(y_t | s_t, \mu, \theta, Y_{t-1}, S_{t-1}) f(s_t | \eta, Y_{t-1}, S_{t-1}). \quad (3.1)$$

The conditional density of y_t is the Gaussian density

$$f(y_t | s_t, \mu, \theta, Y_{t-1}, S_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right), \quad (3.2)$$

where σ_t^2 , defined by equation (2.4), is a function of θ . The marginal density (or probability mass function) of s_t is specified by

$$f(s_t | \eta, Y_{t-1}, S_{t-1}) = f(s_t | \eta, s_{t-1}) = \eta_{s_t s_{t-1}} \quad (3.3)$$

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 $y = (y_1, y_2, \dots, y_T)$ and $S = (s_1, s_2, \dots, s_T)$ given the parameters is then obtained by taking the product of the densities in (3.2) and (3.3) over all observations:

$$f(y, S | \mu, \theta, \eta) \propto \prod_{t=1}^T \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right) \eta_{s_t s_{t-1}}. \quad (3.4)$$

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. The purpose of this algorithm is to simulate draws from the posterior density. These draws then serve to estimate features of the posterior distribution, like means, standard deviations and marginal densities. As the posterior density is not standard we cannot sample from it directly. Gibbs sampling is an iterative procedure to sample sequentially from the posterior distribution; see Gelfand and Smith (1990). Each iteration in the Gibbs sampler produces a draw from a Markov chain, implying that draws are dependent. Under regularity conditions, see e.g. Robert and Casella (2004), the simulated distribution converges to the posterior distribution. Thus a warm-up phase is needed, such that when a sufficient large number of draws has been generated by the Gibbs sampler, the draws that are generated after the warm-up phase can be considered as draws from the posterior distribution. The Markov chain is generated by drawing iteratively from lower dimensional distributions, called blocks or full conditional distributions, of this joint posterior distribution. These full conditional distributions are easier to sample from because either they are known in closed form or simulated by a lower dimensional auxiliary sampler. For the MS-GARCH model, the blocks of parameters are given by (θ, μ) , η and the elements of S . We present below a sketch of the Gibbs algorithm, followed by a detailed presentation of the corresponding full conditional densities in three subsections. We explain what are our prior densities for θ , μ and η in these subsections.

Sketch of Gibbs sampling algorithm: The superscript r on a parameter denotes a draw of the parameter at the r th iteration of the algorithm. For iteration 1, initial values η^0 , θ^0 , μ^0 and s_t^0 for $t = 1, 2, \dots, T$ must be used. One iteration of the algorithm involves three steps:

- (1) Sample sequentially each state variable s_t^r for $t = 1, 2, \dots, T$ given η^{r-1} , μ^{r-1} , θ^{r-1} , s_{t-1}^{r-1} , s_{t+1}^{r-1} : see Subsection 3.1.1.
- (2) Sample the transition probabilities η^r given s_t^r for $t = 1, 2, \dots, T$: see Subsection 3.1.2.
- (3) Sample (θ^r, μ^r) given s_t^r for $t = 1, 2, \dots, T$, and η^r : see Subsection 3.1.3.

These three steps are repeated until the convergence of the Markov chain is achieved, which can be evaluated by convergence diagnostics. These warm-up draws are discarded, and the steps are iterated a large number of times to generate draws from which the desired features (means, variances, quantiles, etc.) of the posterior distribution can be estimated consistently.

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 $(s_{t+1}, s_{t+2}, \dots, s_T)$ (because of path dependence of the conditional variances). The full conditional mass function of state t is

$$\varphi(s_t | S_{\neq t}, \mu, \theta, \eta, y) \propto \eta_{1, s_{t-1}}^{2-s_t} \eta_{2, s_{t-1}}^{s_t-1} \eta_{1, s_t}^{2-s_{t+1}} \eta_{2, s_t}^{s_{t+1}-1} \prod_{j=t}^T \sigma_j^{-1} \exp\left(-\frac{(y_j - \mu_{s_j})^2}{2\sigma_j^2}\right), \quad (3.5)$$

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 (1 or 2) we compute the expression above for each of these values, and divide each evaluation by the sum of the two to get the normalized discrete distribution of s_t from which to sample. Sampling from such a distribution once the probabilities are known is like sampling from a Bernoulli distribution.

3.1.2. *Sampling η .* Given a prior density $\pi(\eta)$,

$$\varphi(\eta|S, \mu, \theta, y) \propto \pi(\eta) \prod_{t=1}^T \eta_{s_t, s_{t-1}}, \quad (3.6)$$

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,

$$\varphi(\eta_{11}|S) \propto \eta_{11}^{a_{11}+n_{11}-1} (1 - \eta_{11})^{a_{21}+n_{21}-1}, \quad (3.7)$$

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 $a_{11} = a_{21} = 1$ and is what we use in our simulations and applications below.

3.1.3. *Sampling θ and μ .* Given a prior density $\pi(\theta, \mu)$,

$$\varphi(\theta, \mu|S, \eta, y) \propto \pi(\theta, \mu) \prod_{t=1}^T \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right), \quad (3.8)$$

which does not depend on η . We can sample (θ, μ) either by a griddy-Gibbs step or by a Metropolis step. The griddy-Gibbs step amounts to sampling by numerical inversion each scalar element of the vector (θ, μ) from its univariate full conditional distribution. Numerical tabulation of the cdf is needed because no full conditional distribution belongs to a known family (such as Gaussian). Thus, this step of the Gibbs algorithm that cycles between the states, η and the remaining parameters can be described as follows at iteration $r + 1$, given draws at iteration r denoted by the superscript (r) attached to the parameters:

- (1) Using (3.8), 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 $S, \beta_1, \alpha_1, \theta_2$, and μ sampled at iteration r , over a grid $(\omega_1^1, \omega_1^2, \dots, \omega_1^G)$, to obtain the vector $G_\kappa = (\kappa_1, \kappa_2, \dots, \kappa_G)$.
- (2) By a deterministic integration rule using M points, compute $G_f = (f_0, f_2, \dots, f_G)$ where

$$f_i = \int_{\omega_1^1}^{\omega_1^i} \kappa(\omega_1|S^{(r)}, \beta_1^{(r)}, \alpha_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y) d\omega_1, \quad i = 2, \dots, G. \quad (3.9)$$

- (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 and μ 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, and in practice we choose independent uniform densities for all elements of (θ, μ) .

For the Metropolis version of this block of the Gibbs algorithm, we construct a multivariate Gaussian proposal density for (θ, μ) . Its mean and variance–covariance matrix are renewed at each iteration of the Gibbs algorithm in order to account for the updating of the states sampled as described in Subsection 3.1.1. Thus, after updating the states, we maximize the likelihood of

y given the sampled states, defined by the product of (3.2) over all observations. The mean of the Gaussian proposal is set equal to the ML estimate and the variance–covariance matrix to minus 1 times the inverse-Hessian evaluated at the ML estimate.

3.2. Identification and prior density

In Markov-switching models similar to the model of this paper, one must use some identification restrictions to avoid the label switching issue. This would occur when the states and the parameters can be permuted without changing the posterior distribution. There are different ways to avoid label switching; see Hamilton et al. (2007) for a discussion. In an ML set-up, an identification restriction can be e.g. that the mean of regime 1 (μ_1) is larger than the mean of regime 2 (μ_2), or the same restriction for the variance level (ω) or another parameter of one GARCH component relative to the other. These restrictions aim at avoiding confusion between the regimes. They are exact restrictions, holding with probability equal to 1 if we reason in the Bayesian paradigm. In Bayesian inference, the restrictions need not be imposed with probability equal to 1. They can be imposed less stringently through the prior density. For example, coming back to the restriction on means, the prior for μ_1 can be uniform on $(-0.03, +0.09)$ and independent of the prior for μ_2 taken as uniform on $(-0.09, +0.03)$. These priors overlap but not too much, allowing a sufficiently clear *a priori* separation of the two regimes and avoiding label switching in the MCMC posterior simulation.

Thus the regimes must be sufficiently separated to be identified; that is, some parameters must be different between regimes. Our approach to this is to use prior supports for the corresponding parameters of the two regimes that are partially different. When choosing these prior supports, we must be careful to avoid two problems: truncating the posterior density (if we use too narrow supports compared to the likelihood location), or computing the posterior over too wide a region of the parameter space (if we use too wide prior supports). In the former case, the prior density will distort the sample information strongly and thus the posterior results will not be interesting. In the latter case, the posterior computations will be inefficient because the prior will be multiplied by a quasi-null likelihood value on a big portion of the parameter space.

Thus our approach is to start with wide enough prior intervals, taking care of the need to separate the regimes, and to narrow these intervals if possible or to widen them if needed, such that finally the posterior is not significantly truncated and label switching is avoided. If an ML estimation would be feasible, the analogue procedure would be to impose exactly some identification constraints and to choose the starting values for the maximization in a portion of the parameter space where the likelihood value is not very small.

3.3. Extensions

Although we presented the Gibbs sampling algorithm for the case of two states, some extensions are possible without changing the nature of the algorithm, but they will increase the computation time. A first extension consists of allowing the mean of the process to switch between two ARMA processes rather than constant means. Similarly, one can consider GARCH(p, q) specifications for the regimes with p and q larger than 1. Such extensions are dealt with by redefining the θ and μ parameter vectors and adapting directly the procedures described in Subsection 3.1.3 to account for the additional parameters. Henneke et al. (2006) describe a slightly different MCMC algorithm from ours for this more general model, considering also the case of a Student distribution for the innovation.

A second extension consists of considering more than two regimes. Again, the algorithm described for two regimes can in principle be extended. The states must be sampled from a discrete distribution with three points of support, and the η parameters from Dirichlet distributions that generalize Beta distributions and can be simulated easily. Finally, the nature of the third block does not change and can be sampled as for two regimes. In practice, an increase in the number of regimes increases the number of parameters, which is especially costly in computing time for the third block (θ and μ), whether one uses a griddy-Gibbs or a Metropolis step. A related difficulty lies in the identification of the regimes; see the discussion in the previous subsection. Specifying the prior density for more than two regimes requires more search and it is obvious that in this case the algorithm will be more complicated. We leave the detailed study of this type of question for further research.

Another issue that is not yet solved is the computation of the marginal likelihood of the MS-GARCH model, as recognized in the literature; see e.g. Maheu and He (2009). Because of the combination of the GARCH (rather than ARCH) structure and the latent variable structure, we could not find a method to compute the marginal likelihood. Chib (1996) has developed an algorithm for Markov mixture models. Kaufmann and Frühwirth-Schnatter (2002) use this algorithm for an MS-ARCH model, and mention that it cannot be extended to the MS-GARCH case because of the path dependence problem. Equation (3) in Chib (1996) does not hold in case of path dependence. It would be quite a valuable contribution to solve this issue since it would allow us to do model selection in a fully Bayesian way. Short of this, we resort to the following procedure: once the Gibbs sampler has been applied, we compute 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. With these means, we can assign each observation to one regime, by attributing an observation to the regime for which the state has the highest posterior probability. Once the data are classified in this way, we can compute easily the Bayesian information criterion (BIC) using the likelihood function defined by the product over all observations of the contributions given by the densities in (3.2). Given the values of the state variables, it is easy to evaluate this function (in logarithm) at the posterior means of the parameters and subtract from it the usual penalty term $0.5p \log T$, where p is the number of parameters, to get the BIC value. In large samples, the BIC usually leads to choose the model also picked by the marginal likelihood criterion; see the discussion in Kass and Raftery (1995). We use this method in the application (Section 4).

3.4. Monte Carlo study

We have simulated a data-generating process (DGP) corresponding to the model defined by equations (2.3)–(2.4) 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 summary statistics for 1500 observations from this DGP, and in Figure 3 we show the series, its estimated density and the autocorrelations of the squared data. The mean of the data is close to zero. The density is skewed to the left, and its excess kurtosis is estimated to be 5.52 (the excess kurtosis is 1.62 for the first component GARCH and 0.12 for the second). The ACF of

Table 1. Descriptive statistics for the simulated data.

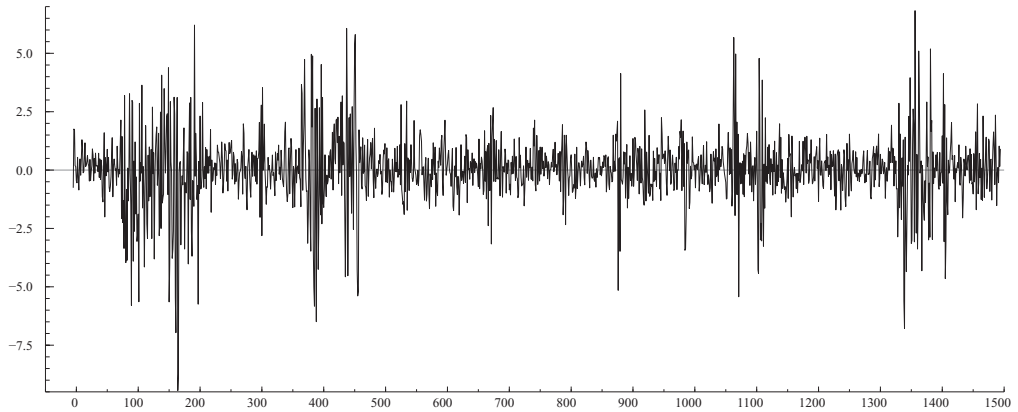
Mean	0.010	Maximum	6.83
Standard deviation	1.472	Minimum	−9.44
Skewness	−0.616	Kurtosis	8.52

Note: Statistics for 1500 observations of the DGP defined in Table 2.

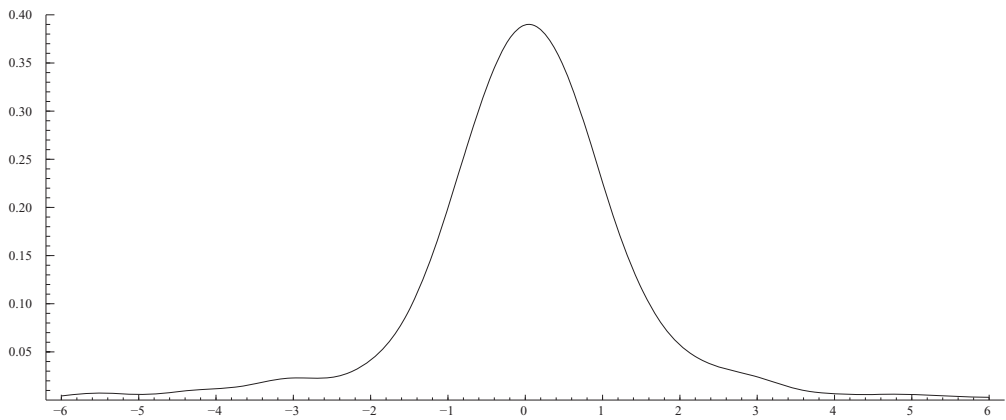
the squared data is strikingly more persistent than the ACF of each GARCH component, which are both virtually at 0 after 10 lags. Said differently, a GARCH(1,1) process would have to be close to integrated to produce the excess kurtosis and the ACF shown in Figure 3. Thus we can say that the DGP chosen for illustrating the algorithm is representative of a realistic process for daily financial returns.

In Table 2, we report the posterior means and standard deviations for the model corresponding to the DGP, using the simulated data described above. The results are in the last two columns of the table. In Figure 4, 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 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 bimodalities 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% of the data are correctly classified.

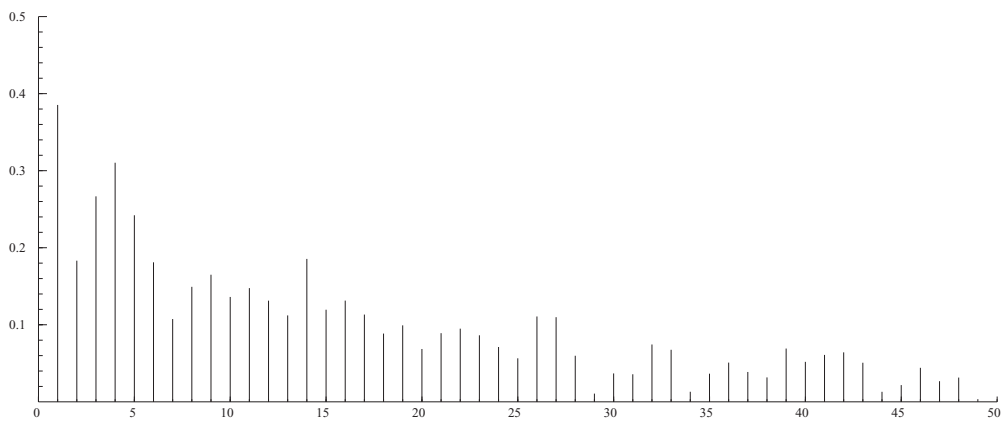
We repeated the previous experiment 100 times, thus generating 100 samples of size 1500 of the same DGP and repeating the Bayesian estimation for each sample, using the same prior throughout these 100 repetitions. In columns 4 and 5 of Table 2, we report the means and standard deviations of the 100 posterior means of each parameter. The posterior mean of each parameter can be viewed as an estimator (i.e. a function of the sample) of the parameter, and the difference between the mean of these posterior means and the true value measures the bias of this estimator, while the standard deviation of these means is a measure of the sampling variability of the estimator. The reported results show that the bias is not large in general, and that it is larger for the parameters of the second regime than of the first. The standard deviations are also larger for the parameters of the second regime. These larger biases and standard deviations are not surprising since the second regime is less frequently active than the first one. The interpretation of the posterior mean as an estimator that is not much biased (for samples of 1500 observations) is of course also conditional on the prior information we have put in the estimation (a non-informative prior on finite intervals). It should be kept in mind that by changing the location and precision of the prior, one could induce much more (or less) bias and increase (or decrease) the variability of the estimator. From a strictly Bayesian viewpoint, the issue of bias and sampling variability is not relevant. Nevertheless, we can safely conclude from these 100 repetitions that



(a) Sample path (1500 observations)



(b) Kernel density



(c) Correlogram of squared data

Figure 3. Graphs for simulated data for DGP defined in Table 2.

Table 2. Monte Carlo results.

			100 repetitions		Single sample	
	DGP values	Prior supports	Mean	Std. dev.	Mean	Std. dev.
ω_1	0.30	(0.15 0.45)	0.301	(0.043)	0.345	(0.051)
β_1	0.20	(0.05 0.40)	0.201	(0.061)	0.192	(0.087)
α_1	0.35	(0.10 0.50)	0.355	(0.059)	0.264	(0.051)
ω_2	2.00	(0.50 4.00)	2.232	(0.513)	2.136	(0.688)
β_2	0.60	(0.35 0.85)	0.556	(0.084)	0.584	(0.106)
α_2	0.10	(0.02 0.35)	0.110	(0.043)	0.142	(0.049)
μ_1	0.06	(0.02 0.15)	0.056	(0.017)	0.079	(0.016)
μ_2	-0.09	(-0.35 0.18)	-0.056	(0.085)	-0.076	(0.103)
η_{11}	0.98	(0.00 1.00)	0.977	(0.005)	0.987	(0.004)
η_{22}	0.96	(0.00 1.00)	0.951	(0.016)	0.958	(0.012)

Note: Columns 4–5: Mean and standard deviations of posterior means for the MS-GARCH model. Results based on 100 replications, each of which consists of 1500 observations from the DGP defined by equations (2.3)–(2.4) with $N(0, 1)$ distribution. Columns 6–7: Posterior means and standard deviations for a single sample of 1500 observations.

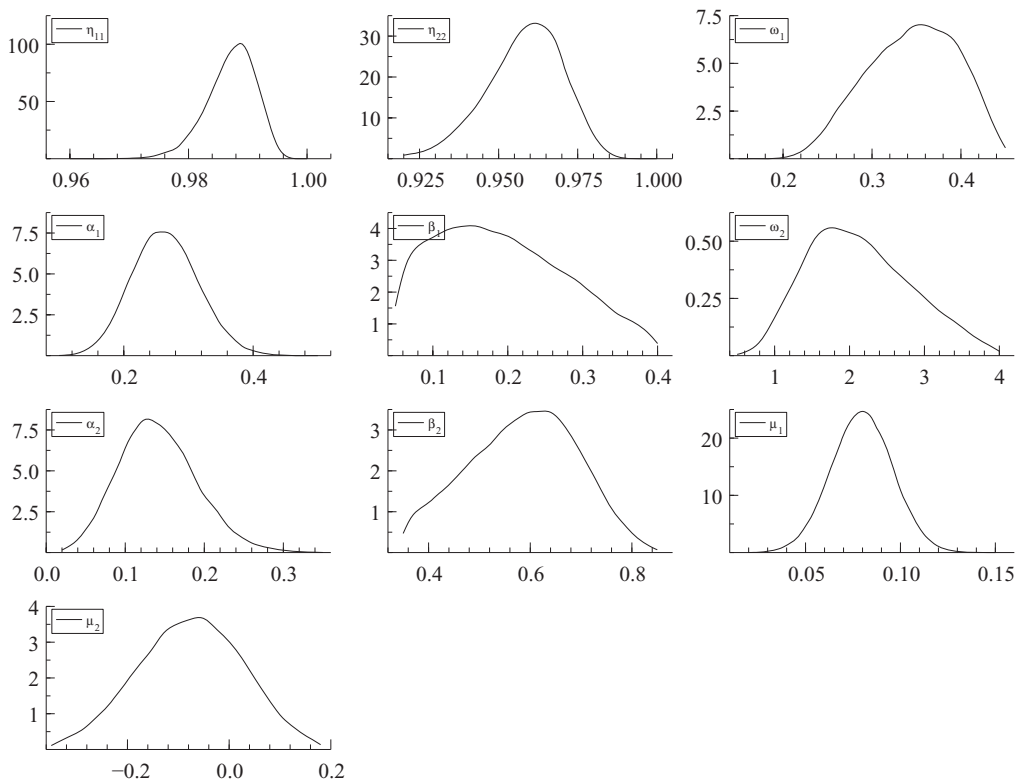


Figure 4. Posterior densities for the MS-GARCH model.

indeed our MCMC algorithm is able to estimate the MS-GARCH model in a reliable way, since for 100 repetitions it was able to infer the DGP parameters from the data quite satisfactorily.

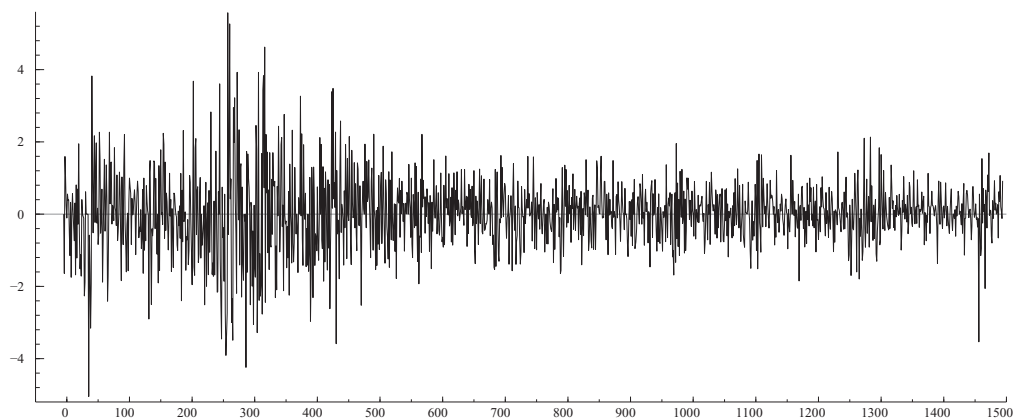
Finally, let us mention that we also simulated a three-state model by adding a third, more persistent regime, to those reported in Table 2. The algorithm is also able to recover the DGP. The detailed results are available on request.

4. APPLICATION

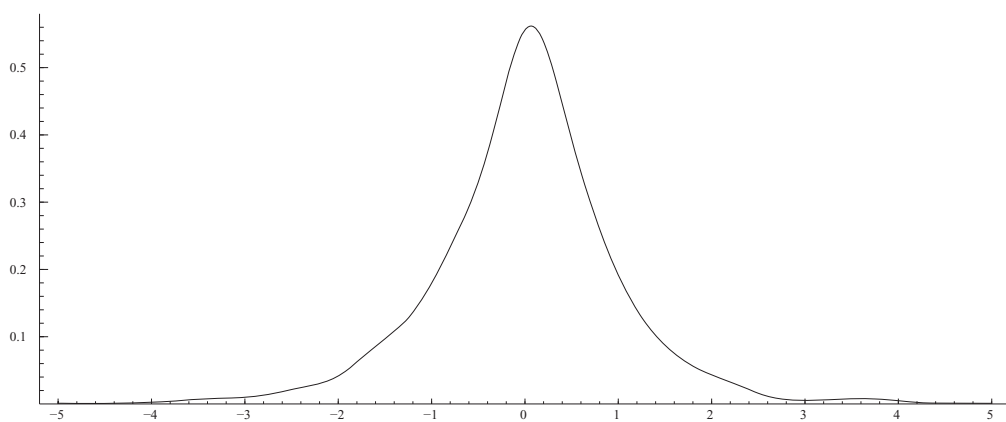
We use the S&P500 daily percentage returns from 19/07/2001 to 20/04/2007 (1500 observations) for estimation. Figure 5 displays the sample path, the kernel density and the correlogram of the squared returns. We observe a strong persistence in the squared returns, a slightly positive skewness and a usual excess kurtosis for this type of data and sample size; see also Table 3.

In Table 4, we report the posterior means and standard deviations from the estimation of two models using the estimation sample. The estimated models are the two-regime MS-ARCH model defined by setting $\beta_1 = \beta_2 = 0$ in equations (2.3)–(2.4), 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 6 and 7. The intervals over which the densities are drawn are the prior intervals (except for the transition probabilities). The intervals for the GARCH parameters were chosen to avoid negative values and truncation. For the MS-GARCH model, we report in the table the results obtained by using the two versions of the Gibbs algorithm described in Subsection 3.1.3, one (GG) using the griddy-Gibbs method for sampling the mean and variance equation parameters, and the other (MH) using a Metropolis–Hastings step for the same parameters. In all cases, the total Gibbs sample size is equal to 50,000 observations with a warm-up sample of 20,000, and the prior distribution is the same. The MH version needs 20% more computing time than the griddy-Gibbs one, and its acceptance rate is 68%, which is a good performance. However, the number of rejections due to the prior restrictions (finite ranges of the GARCH parameters) is slightly more than 50%. These rejections are not needed when using the griddy-Gibbs version. Thus we may conclude that the GG version has a slight advantage in this instance, but more importantly, it is reassuring that both algorithms give approximately the same posterior results.

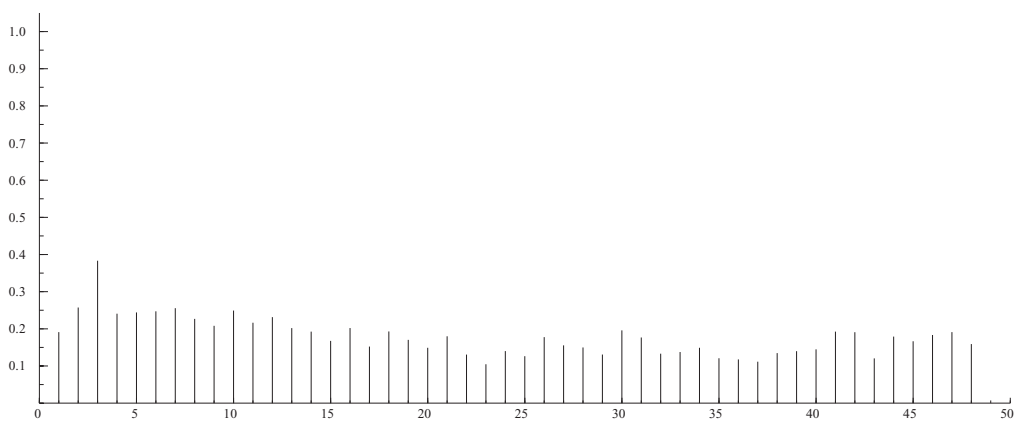
When estimating the MS-ARCH model, we find that in the first regime, which is characterized by a low volatility level ($\omega_1/(1 - \alpha_1) = 0.42$ using the posterior means as estimates, as opposed to 2.24 in the second regime), the ARCH coefficient α_1 is close to 0 (posterior mean 0.014, standard deviation 0.012; see also the marginal density in Figure 6). This is a weak evidence in favour of a dynamic 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 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 $\omega_2/(1 - \alpha_2 - \beta_2) = 1.73$.



(a) Sample path (1500 observations)



(b) Kernel density



(c) Correlogram of squared data

Figure 5. Graphs for S&P500 daily returns from 19/07/2001 to 20/04/2007.

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

Note: Sample period: 19/07/2001 to 20/04/2007 (1500 observations).

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

	MS-ARCH (GG)		MS-GARCH (GG)		MS-GARCH (MH)	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
ω_1	0.419	(0.028)	0.308	(0.025)	0.313	(0.027)
β_1	–		–		–	
α_1	0.014	(0.012)	–		–	
ω_2	1.988	(0.175)	0.0467	(0.024)	0.0486	(0.025)
β_2	–		0.919	(0.024)	0.917	(0.021)
α_2	0.115	(0.042)	0.054	(0.015)	0.055	(0.012)
μ_1	0.046	(0.016)	0.071	(0.019)	0.069	(0.020)
μ_2	−0.040	(0.044)	−0.012	(0.029)	−0.012	(0.032)
η_{11}	0.994	(0.003)	0.978	(0.011)	0.979	(0.012)
η_{22}	0.986	(0.006)	0.985	(0.006)	0.986	(0.006)
π_1	0.300		0.595		0.600	

Note: Sample period: 19/07/2001 to 20/04/2007 (1500 observations). A – symbol means that the parameter was set to 0. GG: gridy-Gibbs, MH: Metropolis–Hastings. π_1 : unconditional probability of state 1.

Next, we compare the two models, starting with their BIC values, which are computed as explained at the end of Subsection 3.3. These values show clearly that the MS-GARCH model is strongly favoured, with a BIC value of −954.72, compared to the value of −1034.2 for the MS-ARCH.

Another way to compare the 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 8 and 9 display the paths of these means. Both figures show, in conjunction with the sample path of the data (in Figure 5), that high probabilities are associated with high volatility periods (observations 1 to 500 and some peaks later, which are less strong for the ARCH than the GARCH model). From the posterior means of the MS-GARCH model, we can also deduce that the unconditional probabilities of the regimes are, respectively, 0.60 ($= (1 - \eta_{11}) / (2 - \eta_{11} - \eta_{22})$) for the first one and 0.40 for the second one. For the MS-ARCH model, the steady-state probability of regime 1 is 0.30 only. These proportions correspond intuitively to the information provided by the plots of the mean states.

Finally, we provide some diagnostics about the estimated models. To compute standardized residuals, we attribute an observation to regime 2 if the mean of the corresponding state variable is larger than 0.5 and to regime 1 otherwise, then we standardize the observation by subtracting its mean and dividing by its conditional standard deviation computed for the attributed state

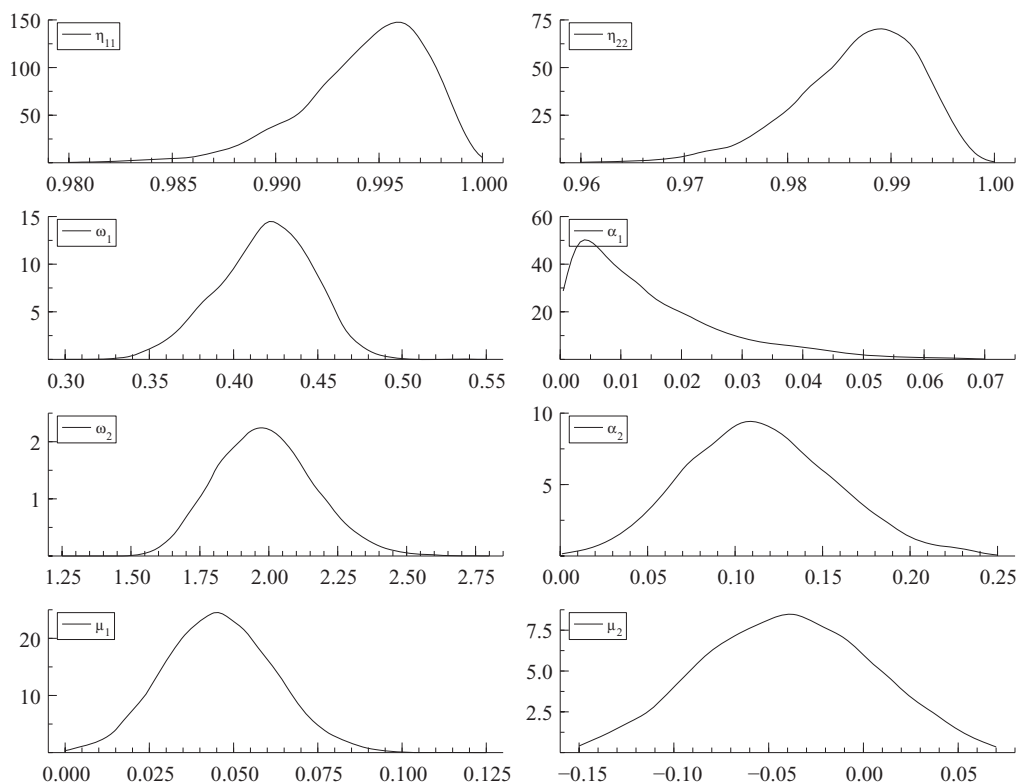


Figure 6. Posterior densities for the MS-ARCH model (S&P500 daily returns).

using the posterior means. The Ljung–Box $Q(p)$ -statistics of the residuals of the MS-ARCH model are $Q(10) = 8.19$, $Q(20) = 14.8$ for the residuals and $Q(10) = 83.5$, $Q(20) = 110$ for the squared residuals. For the MS-GARCH model, the corresponding statistics are $Q(10) = 10.1$, $Q(20) = 16.4$ for the residuals and $Q(10) = 16.5$, $Q(20) = 22.1$ for the squared residuals. Even if there is no formal statistical theory that allows us to compare these values to quantiles of chi-square distributions, the informal evidence from these results is that there is clearly some misspecification of the conditional variance dynamics in the MS-ARCH model, and not in the MS-GARCH one. The conditional mean of both models seems correctly specified.

5. CONCLUSION

We establish some theoretical properties of a Markov-switching GARCH model with constant transition probabilities. We provide simple sufficient conditions for the ergodic stationarity of the process and the existence of its moments. Since ML estimation is not feasible due to path dependence, we develop a reliable Bayesian estimation algorithm for this model, and we apply it to a sample of S&P500 daily returns. Based on residual diagnostics and the Bayesian information

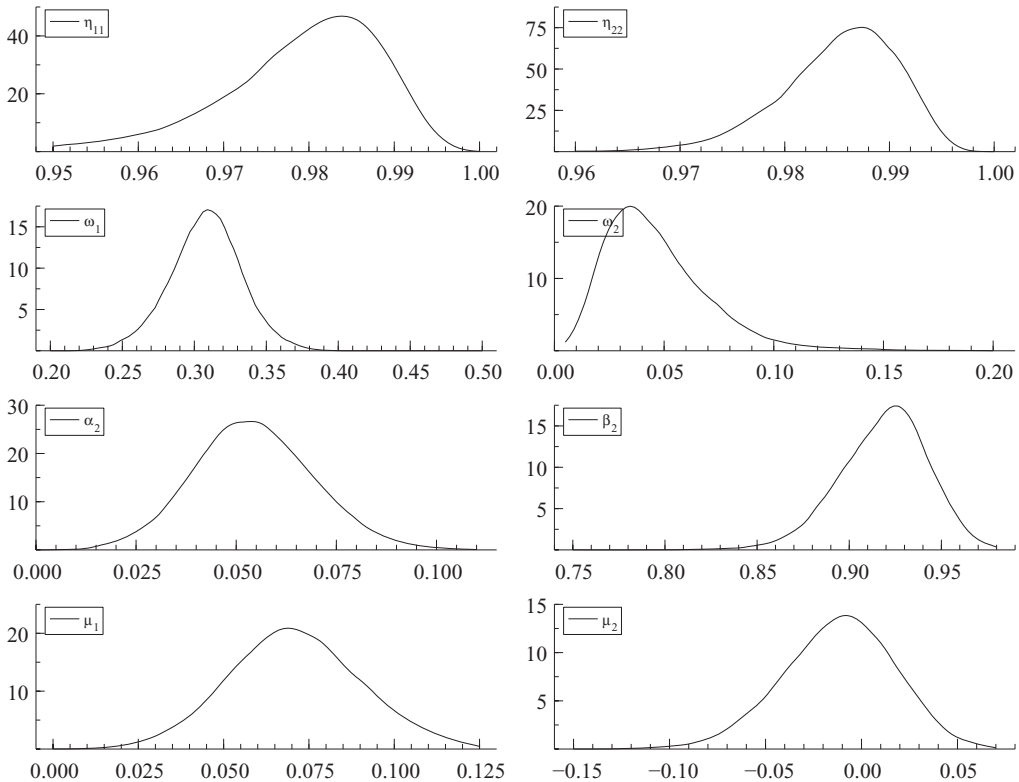


Figure 7. Posterior densities for the MS-GARCH model (S&P500 daily returns).

criterion we find that the MS-GARCH model with two regimes fits the data much better than the corresponding ARCH model.

Further research could be oriented in several directions. A first direction could be to refine the specification by using existing extensions of the simple GARCH(1,1) model, and allowing an ARMA structure for the conditional mean. A second direction of research is to specify the transition probabilities as a function of past information as in Gray (1996). All these extensions would render the algorithm more CPU-time consuming (due to the additional parameters) but would not complicate it fundamentally. Establishing the geometric ergodicity and existence of moments of such more richly specified processes would require us to extend and adapt the proofs presented in the current paper. A second topic would be to compare our MS-GARCH model to other models like Gray (1996), the MS and finite mixture GARCH models of Haas et al. (2004) and rank their performance in terms of in-sample fit and out-of-sample forecasting of volatility. Finally, further research could be focused on estimating the model with other data series, more regimes and in comparisons with other GARCH models, in a similar way as done by Marcucci (2005). An open issue of particular relevance would be to find a way to compute the marginal likelihood of the MS-GARCH model.

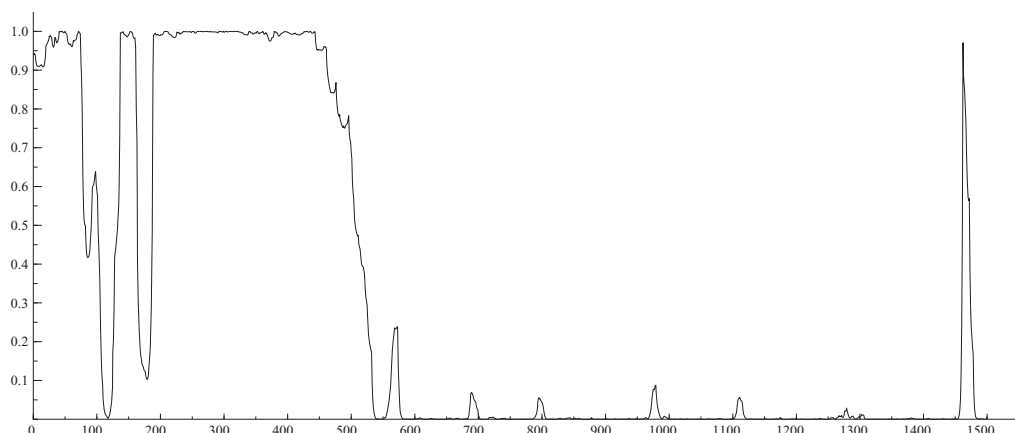


Figure 8. Mean states MS-ARCH.

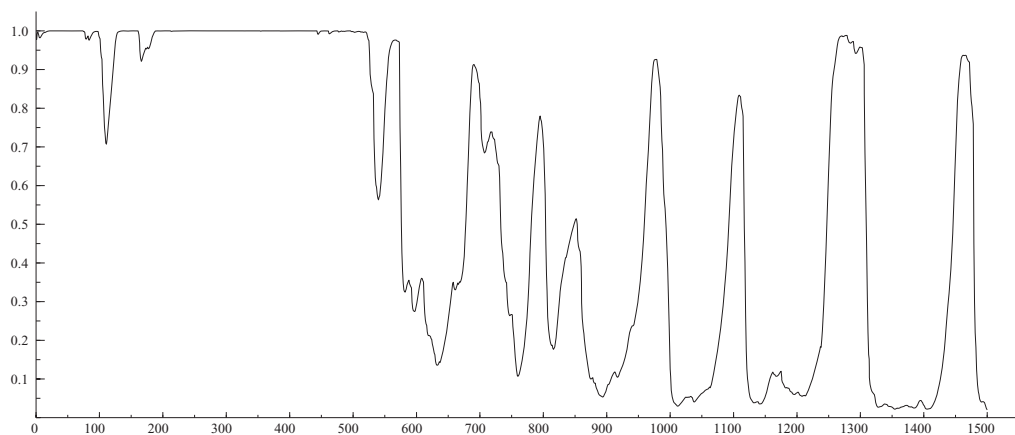


Figure 9. Mean states MS-GARCH.

ACKNOWLEDGMENTS

We thank the referees and the editor for their insightful comments on an earlier version. Bauwens's work was supported by the contract 'Projet d'Actions de Recherche Concertées' 07/12-002 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'. Luc Bauwens is also a member of ECORE, an association between ECARES and CORE. 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 paper 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.

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 Discussion Paper 2006/11, Université catholique de Louvain, Louvain La Neuve.
- Bauwens, L. and J. Rombouts (2004). Econometrics. In J. Gentle, W. Hardle and Y. Mori (Eds.), *Handbook of Computational Statistics: Concepts and Methods*, 951–79. Heidelberg: Springer.
- Bollen, N., S. Gray and R. Whaley (2000). Regime-switching in foreign exchange rates: evidence from currency option prices. *Journal of Econometrics* 94, 239–76.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–27.
- Bollerslev, T., R. F. Engle and D. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, 2959–3038. Amsterdam: North Holland.
- Cai, J. (1994). Markov model of unconditional variance in ARCH. *Journal of Business and Economics Statistics* 12, 309–16.
- Chan, K. (1993). A review of some limit theorems of Markov chains and their applications. In H. Tong (Ed.), *Dimension Estimation and Models*. River Edge, NJ: World Scientific Publishing.
- Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* 75, 79–97.
- Das, D. and B. H. Yoo (2004). A Bayesian MCMC algorithm for Markov switching GARCH models. Working paper, City University of New York and Rutgers University.
- Davidson, J. (1994). *Stochastic Limit Theory*. New York: Oxford University Press.
- Diebold, F. (1986). Comment on modelling the persistence of conditional variances. *Econometric Reviews* 5, 51–56.
- Diebold, F. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics* 105, 131–59.
- Douc, R., E. Moulines and T. Ryden (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Annals of Statistics* 32, 2254–304.
- Doukhan, P. (1994). *Mixing: Properties and Examples*. New York: Springer.
- 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, 815–18.
- Francq, C. and J.-M. Zakoian (2005). The l^2 -structures of standard and switching-regime GARCH models. *Stochastic Processes and their Applications* 115, 1557–82.
- Gelfand, A. and A. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Gerlach, R. and F. Tuyl (2006). MCMC methods for comparing stochastic volatility and GARCH models. *International Journal of Forecasting* 22, 91–107.
- Giraitis, L., R. Leipus and D. Surgailis (2007). Recent advances in ARCH modelling. In G. Teyssière and A. Kirman (Eds.), *Long Memory in Economics*, 3–38. Heidelberg: Springer.
- 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. Paoletta (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, 307–33.
- Hamilton, J., D. Waggoner and T. Zha (2007). Normalization in econometrics. *Econometric Reviews* 26, 221–52.
- 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.
- Hilebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics* 129, 121–38.
- Jacquier, E., N. Polson and P. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* 12, 371–417.
- Kass, R. and A. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773–95.
- Kaufman, S. and S. Frühwirth-Schnatter (2002). Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis* 23, 425–58.
- Kaufman, S. and M. Scheicher (2006). A switching ARCH model for the German DAX index. *Studies in Nonlinear Dynamics and Econometrics* 10, No. 4, Article 3.
- Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27, 363–94.
- Kramer, W. (2008). Long memory with Markov-switching GARCH. *Economics Letters* 99, 390–92.
- Lamoureux, C. and W. Lastrapes (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economic Statistics* 8, 225–34.
- Lutkepohl, H. (1996). *Handbook of Matrices*. New York: John Wiley.
- Maheu, J. and Z. He (2009). Real time detection of structural breaks in GARCH models. Working paper, Department of Economics, University of Toronto.
- Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 9, 1–53.
- Meitz, S. and P. Saikkonen (2008). Stability of nonlinear AR-GARCH models. *Journal of Time Series Analysis* 29, 453–75.
- Meyn, S. and R. Tweedie (1993). *Markov Chains and Stochastic Stability*. London: Springer.
- 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, 378–90.
- Moeenaddin, R. and H. Tong (1990). Numerical evaluations of distribution of non-linear autoregression. *Journal of Time Series Analysis* 11, 33–48.
- Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–34.
- Robert, C. and G. Casella (2004). *Monte Carlo Statistical Methods*. New York: Springer.
- Schwert, G. (1989). Why does stock market volatility change over time? *Journal of Finance* 44, 1115–53.
- Stachurski, J. and V. Martin (2008). Computing the distributions of economic models via simulations. *Econometrica* 76, 443–50.
- Tanner, M. and W. Wong (1987). The calculation of the posterior distributions by data augmentation. *Journal of the American Statistical Association* 82, 528–40.
- Tjostheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.
- Tsay, R. (2005). *Analysis of Financial Time Series*. New York: John Wiley.
- 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, 265–70.
- 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.

APPENDIX A: PROOFS OF RESULTS

To prove Theorems 2.1 and 2.2, 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 ν denote the Lebesgue and the counting measures, respectively.

Proof of Theorem 2.1: There exists a measurable function $g : S \times \mathfrak{N} \rightarrow S$ 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 $\bar{s}_t = s_{t+1}$, $\eta_t = (u_t, \xi_t)'$, and Z_t be defined on $D \subset \mathfrak{N} \times \mathfrak{N}_+ \times S$ where $\mathfrak{N}_+ = (0, +\infty)$. From (2.3) and (2.4), we have

$$\begin{aligned} Z_t &= \begin{pmatrix} y_t \\ h_t \\ \bar{s}_t \end{pmatrix} = \begin{pmatrix} \mu_{\bar{s}_{t-1}} + \sqrt{h_{t-1}}u_t \\ \omega_{g(\bar{s}_{t-1}, \xi_{t+1})} + \alpha_{g(\bar{s}_{t-1}, \xi_{t+1})}\varepsilon_t^2 + \beta_{g(\bar{s}_{t-1}, \xi_{t+1})}h_{t-1} \\ g(\bar{s}_{t-1}, \xi_{t+1}) \end{pmatrix} \\ &= \begin{pmatrix} \mu_{\bar{s}_{t-1}} + \sqrt{h_{t-1}}u_t \\ \omega_{g(\bar{s}_{t-1}, \xi_{t+1})} + (\alpha_{g(\bar{s}_{t-1}, \xi_{t+1})}u_t^2 + \beta_{g(\bar{s}_{t-1}, \xi_{t+1})})h_{t-1} \\ g(\bar{s}_{t-1}, \xi_{t+1}) \end{pmatrix} = F(Z_{t-1}, \eta_t), \end{aligned} \quad (\text{A.1})$$

where $F : D \times \mathfrak{N}^2 \rightarrow D$. Since η_t is independent of Z_{t-1} it follows from (A.1) that $(y_t, h_t, \bar{s}_t)'$ forms a homogeneous Markov chain.

The process is defined on $(D, \mathfrak{F}, \varphi)$. The state space of the process is given by $D = \{(y, h, s) \in \mathfrak{N} \times \mathfrak{N}_+ \times S : (y, h) \in \bigcup_{i=1}^n D_i, s \in S\}$, where D_i is the domain of the chain in each regime and is given by $D_i = \bigcup_{j=1}^n \{(y, h) \in \mathfrak{N} \times \mathfrak{N}_+ : h \geq \omega_i + \alpha_i(y - \mu_j)^2 + \beta_i \bar{h}\}$ and $\bar{h} = \min_{\beta_i < 1} \{\omega_i / (1 - \beta_i)\}$. 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. Note that \bar{h} is a lower bound for the volatility, which implies that $h_t \geq \omega_i + \alpha_i(y_t - \mu_j)^2 + \beta_i \bar{h}$ a.s. The subsets in $\mathfrak{N} \times \mathfrak{N}_+ \times S$ which are not contained in D are transient sets and are null with respect to π , the stationary measure of the process, if it exists (see Chan, 1993). Therefore, without loss of generality our state space excludes such π -null sets (see Zhang et al., 2001). The state space is equipped with \mathfrak{F} , the Borel σ -algebra on $\mathfrak{N} \times \mathfrak{N}_+ \times S$ restricted to D .² The measure φ is the product measure $\lambda^2 \otimes \nu$ on (D, \mathfrak{F}) . 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, \bar{s}_t) moves from (y_0, h_0, \bar{s}_0) to the set $A \in \mathfrak{F}$ in m steps.

In order to establish 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 $k \geq 1$, all $z_0 \in D$ and any Borel measurable sets $A \in \mathfrak{F}$ with positive φ measure (see Chan, 1993). In this case, φ is called an irreducibility measure. Now, since the φ measure of the set of all boundary points is zero, for any non-null set A we can find a close non-null subset A' which is restricted to be interior to the state space. Next, 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 $\bar{s}_0 = i$, $s = \ell$ and \bar{h} is achieved in regime q ; that is, $\bar{h} = \omega_q / (1 - \beta_q)$.

² The topology over D is based on the product topology, where we use the discrete topology over S (for which every subset is open) and the usual topology on \mathfrak{N}^2 .

Let $\tilde{h} = [h - \omega_\ell - \alpha_\ell(y - \mu_q)^2]/\beta_\ell$ and $\varsigma = \tilde{h} - \bar{h}$, since A' is interior set, $\varsigma > 0$. Thus, there exists a positive integer $m = \min(t \geq 1 : \tilde{h} + 0.5\varsigma + 0.5\beta_q^{-t}\varsigma > \omega_q + \beta_q h_0)$, such that the point (y, h, s) can be reached through the following $m + 1$ intermediate steps³ $w = \{(y_t, h_t, \bar{s}_t)\}_{t=1}^{m+1}$ where $\bar{s}_t = q$, $h_t = \tilde{h} + 0.5\varsigma + 0.5\beta_q^{t-(m+1)}\varsigma$, $y_1 = \mu_i + [(h_1 - \omega_q - \beta_q h_0)/\alpha_q]^{0.5}$ and $y_t = \mu_q + [0.5\varsigma(1 - \beta_q)/\alpha_q]^{0.5}$ for $t > 1$. Note that, given $(\bar{s}_{t-1}, \bar{s}_t, y_t, h_{t-1})$ from (14) we have that $h_t = \omega_{\bar{s}_t} + \alpha_{\bar{s}_t}(y_t - \mu_{\bar{s}_{t-1}})^2 + \beta_{\bar{s}_t}h_{t-1}$ which implies by substitution the value of h_t at each step. In the $m + 1$ th step, we get $h_{m+1} = \tilde{h}$ and by setting $s_{m+2} = \ell$ and $y_{m+2} = y$, we get $h_{m+2} = h$. For any A' in the interior of D there exists a small open ball A'' around z ($A'' \subset A'$), such that all the points in it can be reached from z_0 in $m + 2$ steps as in our construction above. This result follows from the definition of m and the continuity of the steps (y_t, h_t) with respect to (y, h) . The $m + 2$ th step transition probability is absolutely continuous with respect to the φ measure. By Assumptions 2.1 and 2.2,

$$p^{m+2}(z_0, z) \geq \prod_{t=0}^{m+1} f((y_{t+1} - \mu_{\bar{s}_t})/h_t^{0.5}) P(\bar{s}_{t+1}|\bar{s}_t) > 0,$$

which implies that $P^{m+2}(z_0, A) \geq P^{m+2}(z_0, A'') > 0$ and hence the chain is φ -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 $A \in \mathfrak{F}$ and $z_0 \in C$,

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

Therefore, $P(z_0, A)$ is minorised 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 (2.4) and the c_r inequality,⁴ we get

$$\begin{aligned} h_t^{1/t} &\leq [\omega_{\bar{s}_t} + (\alpha_{\bar{s}_t} u_t^2 + \beta_{\bar{s}_t}) h_{t-1}]^{1/t} \\ &\leq (\omega_{\bar{s}_t})^{1/t} + (\alpha_{\bar{s}_t} u_t^2 + \beta_{\bar{s}_t})^{1/t} (\omega_{\bar{s}_{t-1}})^{1/t} + [(\alpha_{\bar{s}_t} u_t^2 + \beta_{\bar{s}_t}) (\alpha_{\bar{s}_{t-1}} u_{t-1}^2 + \beta_{\bar{s}_{t-1}})]^{1/t} h_{t-2}^{1/t} \\ &\vdots \\ &\leq \prod_{j=1}^t (\alpha_{\bar{s}_j} u_j^2 + \beta_{\bar{s}_j})^{1/t} h^{1/t} + \sum_{j=1}^{t-1} (\omega_{\bar{s}_{t-j}})^{1/t} \prod_{i=1}^j (\alpha_{\bar{s}_{t-i+1}} u_{t-i+1}^2 + \beta_{\bar{s}_{t-i+1}})^{1/t} + (\omega_{\bar{s}_t})^{1/t}. \end{aligned} \quad (\text{A.2})$$

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

$$\frac{1}{t} \sum_{j=1}^t \log(\alpha_{\bar{s}_j} u_j^2 + \beta_{\bar{s}_j}) \rightarrow \sum_{i=1}^n \pi_i E[\log(\alpha_i u_i^2 + \beta_i)] < 0 \text{ a.s.}; \quad (\text{A.3})$$

see Chan (1993). This result, Assumption 2.1 and the dominated convergence theorem imply that there exists \bar{t} sufficiently large such that $\delta \geq 1/\bar{t} = p$, and for all $\ell \in S$ and $\bar{t} \leq t$,

$$E \left(\prod_{j=1}^t (\alpha_{\bar{s}_j} u_j^2 + \beta_{\bar{s}_j})^p \mid \bar{s}_0 = \ell \right) \leq \gamma < 1. \quad (\text{A.4})$$

³ The integer ‘ m ’ ensures that in the first step the term $(h_1 - \omega_q - \beta_q h_0)$ is positive, so y_1 is well defined in our state space.

⁴ The c_r inequality: For $r > 0$, $E|X + Y|^r \leq c_r(E|X|^r + E|Y|^r)$, where $c_r = 1$ if $r \leq 1$ and $c_r = 2^{r-1}$ if $r > 1$. See Davidson (1994, p. 140).

As a drift function we use $V(z) = 1 + (\bar{\eta}/\Delta)y^{2p} + h^p$, where $\bar{\eta} = \eta - \gamma$, $\Delta = E|u_t^{2p}|$, $p = 1/\bar{\tau}$, 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 (A.2)–(A.4), we find for some $\bar{t} < t$ that

$$E(h_t^p | z_0 = z) \leq E\left(\prod_{j=1}^t (\alpha_{\bar{s}_j} u_j^2 + \beta_{\bar{s}_j})^p \mid \bar{s}_0 = \ell\right) h^p + M \leq M + \gamma h^p,$$

where $M = E(\sum_{j=1}^{t-1} \omega_{\bar{s}_{t-j}}^p \prod_{i=1}^j (\alpha_{\bar{s}_{t-i+1}} u_{t-i+1}^2 + \beta_{\bar{s}_{t-i+1}})^p + \omega_{\bar{s}_t}^p | \bar{s}_0 = \ell)$ and $M < \infty$ by Assumption 2.1. Therefore,

$$\begin{aligned} E(V(z_t) | z_0 = z) &\leq 1 + M + \gamma h^p + \bar{\eta} E(h_{t-1}^p | z_0 = z) E(u_t^{2p}) / \Delta \\ &\leq 2M + 1 + \eta h^p = h^p \left(\frac{2M+1}{h^p} + \eta \right). \end{aligned}$$

Let $M_1 = (2M+1)/(1-\eta)$, $c = M_1^{1/p}(1 + (\Delta/(\bar{\eta}\eta))^{1/p})$ and $\eta' = [(2M+1)/h^p] + \eta$. For $h + y^2 > c$, if $M_1 > h^p$, we get that $\eta' \in (\eta, 1)$, hence $E(V(z_t) | z_0 = z) \leq \eta' h^p < \eta' V(z)$ where the last inequality follows by the definition of the drift function. Otherwise, if $M_1 \leq h^p$, we note that $y^2 > c - h \geq c - M_1^{1/p} = ((\Delta M_1)/(\bar{\eta}\eta))^{1/p}$, hence $M_1 < ((\bar{\eta}\eta)/\Delta)y^{2p}$ and for some $\eta' \in (\eta, 1)$

$$\begin{aligned} E(V(z_t) | z_0 = z) &\leq 2M + 1 + \eta h^p \leq M_1(\eta + 1 - \eta) = M_1 \\ &\leq \eta(\bar{\eta}/\Delta)y^{2p} < \eta' V(z). \end{aligned}$$

Since the Lyapounov function above is bounded on compact sets we have that $E(V(z_t) | z_0 = z) \leq \eta' V(z) + a \cdot 1_C(z)$ for some $a < \infty$ and for all $z_0 \in D$, 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 geometric ergodic. The finiteness of $E(|y_t|^{2p})$ with respect to the stationary measure follows from Meitz and Saikkonen (2008, Lemma 6). If the process is initiated from its stationary distribution, it further follows that the process is α -mixing with geometrically decaying mixing numbers; see Doukhan (1994, p. 89). \square

Proof of Theorem 2.2: Let I_ℓ be a $1 \times n$ vector that contains 1 on the ℓ th position and zeros elsewhere and $l = (1, \dots, 1)$ is an $n \times 1$ vector. The matrix Ω is positive definite, hence the spectral radius is real and positive and Assumption 2.5 implies that there exists a positive integer \bar{t} such that for all $\bar{t} \leq t$ each element of Ω^t is smaller than $1/n$; that is, $(\Omega^t)_{ij} < 1/n$ (see Lutkepohl, 1996, p. 76, 3(a)). Hence

$$E\left(\prod_{j=1}^t (\alpha_{\bar{s}_j} u_j^2 + \beta_{\bar{s}_j})^k \mid \bar{s}_0 = \ell\right) = (I_\ell \Omega) \Omega^{t-1} \leq \gamma < 1. \quad (\text{A.5})$$

By solving (2.4) recursively and setting $h_0 = h$, we get

$$\begin{aligned} h_t &= [\omega_{\bar{s}_t} + (\alpha_{\bar{s}_t} u_t^2 + \beta_{\bar{s}_t}) h_{t-1}] \\ &= \prod_{j=1}^t (\alpha_{\bar{s}_j} u_j^2 + \beta_{\bar{s}_j}) h + \sum_{j=1}^{t-1} (\omega_{\bar{s}_{t-j}}) \prod_{i=1}^j (\alpha_{\bar{s}_{t-i+1}} u_{t-i+1}^2 + \beta_{\bar{s}_{t-i+1}}) + \omega_{\bar{s}_t}. \end{aligned} \quad (\text{A.6})$$

Now, let $\Delta = E|u_t^{2k}|$ and $\bar{\eta} = \eta - \gamma$ where η is some positive number which satisfies $\gamma < \eta < 1$. We select a drift function of the form $V(y, h, s) = 1 + (\bar{\eta}/\Delta)y^{2k} + h^k$ and a test set $C = \{(y, h, s) \in D : h + y^2 \leq$

$c, s \in S\}$. By the binomial theorem, (A.6), Assumption 2.4 and some tedious calculations, we can find for some $\bar{t} < t$, finite positive constants $\{a_m\}$, $m = 1, \dots, k$, independent of h , such that

$$\begin{aligned} E(V(y_t, h_t) | z_0 = z) &= 1 + E(h_t^k | z_0 = z) + \bar{\eta} E(y_t^{2k} | z_0 = z) / \Delta \\ &\leq 1 + \left[E \left(\prod_{j=1}^t (\alpha_{s_j} u_j^2 + \beta_{s_j})^k \right) + \bar{\eta} \right] h^k + \sum_{m=1}^k a_m h^{k-m} \leq h^k \left(\eta + \sum_{m=1}^k a_m h^{-m} \right). \end{aligned}$$

Next, by applying the same arguments as in Theorem 2.1, the desired result follows. \square