The Econometrics Journal

Econometrics Journal (2009), volume **12**, pp. 105–126.
doi: 10.1111/j.1368-423X.2008.00253.x

Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with Student-*t* innovations

DAVID ARDIA[†]

[†]*Department of Quantitative Economics, University of Fribourg, Switzerland*
E-mail: david.ardia@unifr.ch

First version received: June 2007; final version accepted: July 2008

Summary A Bayesian estimation of a regime-switching threshold asymmetric GARCH model is proposed. The specification is based on a Markov-switching model with Student-*t* innovations and K separate GJR(1,1) processes whose asymmetries are located at free non-positive threshold parameters. The model aims at determining whether or not: (i) structural breaks are present within the volatility dynamics; (ii) asymmetries (leverage effects) are present, and are different between regimes and (iii) the threshold parameters (locations of bad news) are similar between regimes. A novel MCMC scheme is proposed which allows for a fully automatic Bayesian estimation of the model. The presence of two distinct volatility regimes is shown in an empirical application to the Swiss Market Index log-returns. The posterior results indicate no differences with regards to the asymmetries and their thresholds when comparing highly volatile periods with the milder ones. Comparisons with a single-regime specification indicates a better in-sample fit and a better forecasting performance for the Markov-switching model.

Keywords: *Asymmetry*, *Bayesian*, *GARCH*, *Markov-switching*, *SMI*, *Threshold*.

1. INTRODUCTION

Markov-switching GARCH models (MSGARCH) have received a lot of attention in recent years as they provide an explanation for the high persistence in volatility observed with single-regime GARCH models (see, e.g. Lamoureux and Lastrapes, 1990). Furthermore, MSGARCH models allow for a sudden change in the (unconditional) volatility level which leads to significant improvements in volatility forecasts (see, e.g. Dueker, 1997, Klaassen, 2002, and Marcucci, 2005).

In the framework of MSGARCH models, a hidden Markov sequence $\{s_t\}$ with state space $\{1, \ldots, K\}$ allows for discrete changes in the GARCH parameters. Following the seminal work of Hamilton and Susmel (1994), different parametrizations have been proposed to account for changes in the scedastic function's parameters (see, e.g. Gray, 1996, Dueker, 1997, and Klaassen, 2002). However, these specifications lead to computational difficulties. The evaluation of the likelihood function for a sample of length T requires the integration over all K^T possible paths, rendering the Maximum Likelihood (ML) estimation infeasible. Approximations are thus required to shorten the dependence to the state variable history but these lead to difficulties in the interpretation of the variance dynamics in each regime.

In order to avoid these problems, Haas et al. (2004) hypothesize K separate GARCH(1,1) processes for the conditional variance of the MSGARCH model. The conditional variances at time t can be written in vector form as follows:

$$\begin{pmatrix} h_t^1 \\ \vdots \\ h_t^K \end{pmatrix} \doteq \begin{pmatrix} \alpha_0^1 \\ \vdots \\ \alpha_0^K \end{pmatrix} + \begin{pmatrix} \alpha_1^1 \\ \vdots \\ \alpha_1^K \end{pmatrix} y_{t-1}^2 + \begin{pmatrix} \beta^1 \\ \vdots \\ \beta^K \end{pmatrix} \odot \begin{pmatrix} h_{t-1}^1 \\ \vdots \\ h_{t-1}^K \end{pmatrix}, \tag{1.1}$$

where $\odot$ denotes the element-by-element multiplication. The MSGARCH process $\{y_t\}$ is then simply obtained by setting $y_t = \varepsilon_t (h_t^{s_t})^{1/2}$, where $\{\varepsilon_t\}$ is i.i.d. standard Normal distributed. The parameters α_0^k, α_1^k and β^k are therefore the GARCH(1,1) parameters related to the kth state of the nature. Under this specification, the conditional variance is solely a function of the past data and current state s_t which renders the ML estimation feasible. In addition to its appealing computational aspects, the MSGARCH model of Haas et al. (2004) has conceptual advantages. In effect, one reason for specifying Markov-switching models that allow for different GARCH behaviour in each regime is to capture the differences in the variance dynamics in low- and high-volatility periods. As pointed out by Haas et al. (2004, p. 498), a relatively large value of α_1^k and relatively low values of β^k in high-volatility regimes may indicate a tendency to over-react to news, compared to regular periods, while there is less memory in these sub-processes. This interpretation requires a parametrization of MSGARCH models that implies a clear association between the GARCH parameters within regime k, i.e. α_0^k, α_1^k and β^k and the corresponding $\{h_t^k\}$ process. Specification (1.1) allows for that clear-cut interpretation of the variance dynamics in each regime.

This paper generalizes the model of Haas et al. (2004) in three ways. First, we extend the symmetric GARCH specification to account for asymmetric movements between the conditional variances and the underlying time series. This is achieved by replacing the K separate GARCH(1,1) processes by K separate GJR(1,1) processes. The GJR specification of Glosten et al. (1993) has proved to be effective in single-regime models for reproducing the asymmetric behaviour of the conditional variance, a phenomenon observed especially on equity markets and referred to as the leverage effect (Black 1976).[1] One explanation of this empirical fact is that negative returns raise a firm's financial leverage which increases its risk and therefore its equity volatility. In the context of regime-switching models, this first extension will allow us to determine whether the impact of past negative shocks on the conditional variance is different between the regimes. Second, instead of centering the asymmetric function at zero, as in the original GJR specification, we center it at a free non-positive threshold parameter. Our aim is to test whether it is the sign of the past shock (i.e. leverage effect) or the effect of a *bad news* (i.e. a return's value below a given non-positive level) which influences the conditional variance. In a regime-switching framework, this second extension will allow us to determine if bad news is similar between the different volatility regimes. Finally, we consider Student-t innovations instead of Normal innovations. The use of a Student-t distribution enhances the stability of the states and allows to focus on the conditional variance's behaviour instead of capturing some outliers (Klaassen, 2002). Moreover, the Student-t distribution includes the Normal distribution as the limiting case so we have additional flexibility in the modelling.

[1] Other specifications such as the GQARCH model of Sentana (1995) are also able to reproduce the leverage effect. The GJR model is however more often used in empirical applications.

The Bayesian estimation of MSGARCH models has several advantages over the classical approach. First, proper computational methods based on Markov chain Monte Carlo (MCMC) procedures avoid the common problem of local maxima encountered in ML estimation of this class of models (Ardia, 2008, section 7.7). Second, the exploration of the joint posterior distribution gives a complete picture of the parameter uncertainty and this cannot be achieved via the classical approach (see, e.g. Hamilton and Susmel, 1994). In particular, for the model considered in this paper, the asymptotic distribution theory of the ML estimator is inoperable because of the threshold parameters (Geweke and Terui, 1991, p. 43). Third, constraints on the model parameters can be incorporated through prior specifications. Finally, discrimination between models can be achieved through the calculation of model likelihoods and Bayes factors. In the classical framework, testing the null of K versus K' states is not possible (see, e.g. Frühwirth-Schnatter, 2006, section 4.4).

This paper proposes a novel MCMC scheme to perform the Bayesian estimation of MSGARCH models. Our methodology has the advantage of being fully automatic and thus avoids the time-consuming and difficult task of choosing and tuning a sampling algorithm. Non-expert users who need to run the estimation frequently and/or for a large number of time series should find this new procedure helpful in that regard.

As an application, we fit a single-regime threshold GJR(1,1) model and a two-state Markov-switching threshold GJR(1,1) model to the Swiss Market Index log-returns. We use the random permutation sampler of Frühwirth-Schnatter (2001) to find suitable identification constraints and show the presence of two distinct volatility regimes in the time series. The posterior results indicate no difference between asymmetries and locations of the asymmetry for high versus low volatility periods. By using the Deviance information criterion and by estimating the model likelihoods, we show the in-sample superiority of the regime-switching specification. To test the predictive performance of the models, we run a forecasting analysis based on the Value at Risk and estimate the predictive likelihoods. Both measures indicate the better performance for the Markov-switching model.

The paper proceeds as follows. We set up the model in Section 2. The MCMC scheme is detailed in Section 3. The models are estimated in Section 4 where we assess their goodness-of-fit and test their predictive performance. Section 5 concludes.

2. THE MODEL AND THE PRIORS

A Markov-switching threshold GJR(1,1) model with Student-t innovations may be written as follows:

$$\begin{aligned} y_t = {} & \varepsilon_t (\varrho\, \boldsymbol{e}_t' \boldsymbol{h}_t)^{1/2} \qquad \text{for } t = 1, \ldots, T \\ \varepsilon_t \overset{\text{i.i.d.}}{\sim} {} & \mathcal{S}(0, 1, \nu)\,, \end{aligned} \tag{2.1}$$

where $\boldsymbol{e}_t \doteq (\mathbb{I}\{s_t = 1\} \cdots \mathbb{I}\{s_t = K\})'$; $\mathbb{I}\{\bullet\}$ is the indicator function; the sequence $\{s_t\}$ is assumed to be a stationary, irreducible Markov process with discrete state space $\{1, \ldots, K\}$ and transition matrix $P \doteq [P_{ij}]$ where $P_{ij} \doteq \mathbb{P}(s_{t+1} = j \mid s_t = i)$; $\mathcal{S}(0, 1, \nu)$ denotes the standard Student-t density with ν degrees of freedom; $\varrho \doteq (\nu - 2)/\nu$ is a scaling factor which ensures that the conditional variance is given by $\boldsymbol{e}_t' \boldsymbol{h}_t$. We define the vector of threshold GJR(1,1) conditional variances in (2.1) in a compact form as follows:

$$\boldsymbol{h}_t \doteq \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 y_{t-1}^2 + \boldsymbol{\alpha}_2 \odot \mathbf{I}\{y_{t-1} < \boldsymbol{\tau}\} \odot (\boldsymbol{\tau} - y_{t-1})^2 + \boldsymbol{\beta} \odot \boldsymbol{h}_{t-1}\,, \tag{2.2}$$

where $\boldsymbol{h}_t \doteq (h_t^1 \cdots h_t^K)'$, $\boldsymbol{\alpha}_\bullet \doteq (\alpha_\bullet^1 \cdots \alpha_\bullet^K)'$, $\boldsymbol{\beta} \doteq (\beta^1 \cdots \beta^K)'$, $\boldsymbol{\tau} \doteq (\tau^1 \ldots \tau^K)'$ and $\mathbf{I}\{y_{t-1} < \boldsymbol{\tau}\} \doteq (\mathbb{I}\{y_{t-1} < \tau^1\} \cdots \mathbb{I}\{y_{t-1} < \tau^K\})'$. In order to ensure the positivity of the conditional variance in every regime, we require that $\boldsymbol{\alpha}_0 > \mathbf{0}$ and $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta} \geqslant \mathbf{0}$, where $\mathbf{0}$ is a vector of zeros. Moreover, we require that $\boldsymbol{\tau} \leqslant \mathbf{0}$ to ensures that the conditional variance is minimal when the past return is zero (i.e. when the underlying price remains constant over the last time period). Indeed, a positive threshold would yield inconsistencies such as having a return of zero affecting the volatility by more than a positive return. It is also hard to justify why a positive return would be construed as bad news.[2] Finally, we set $\boldsymbol{h}_0 \doteq \mathbf{0}$ and $y_0 \doteq 0$ for convenience.[3]

Specification (2.2) encompasses the Markov-switching GARCH(1,1) model of Haas et al. (2004) when $\boldsymbol{\alpha}_2 \doteq \mathbf{0}$ and the Markov-switching GJR(1,1) model studied in Ardia, (2008, chap. 7) when $\boldsymbol{\tau} \doteq \mathbf{0}$. Single-regime models are obtained in a straightforward manner by setting $K \doteq 1$. As pointed out in Section 1, expression (2.2) allows to determine whether or not: (i) structural breaks are present within the volatility dynamics (e.g. $\alpha_0^k \neq \alpha_0^{k'}$); (ii) an asymmetric response is present (i.e. $\alpha_2^k > 0$ for at least on k), and is different between the regimes (i.e. $\alpha_2^k \neq \alpha_2^{k'}$) and (iii) locations of bad news are similar between the regimes (i.e. $\tau^k \neq \tau^{k'}$). Moreover, expression (2.2) leads to a smooth news impact curve (Engle and Ng, 1993) at the threshold values and is convenient in the construction of the proposal density for the generation of the thresholds (see the Appendix).

The use of a Student-t instead of a Normal distribution is quite popular in standard single-regime GARCH literature. For regime-switching models, a Student-t distribution might be seen as superfluous since the switching regime can account for large unconditional kurtosis in the data. However, as empirically observed by Klaassen (2002), allowing for Student-t innovations within regimes enhances the stability of the states and allows to focus on the conditional variance's behaviour instead of capturing some outliers. Moreover, the Student-t distribution includes the Normal distribution as the limiting case where the degrees of freedom parameter goes to infinity. We have therefore an additional flexibility in the modelling and can impose Normality by constraining the lower boundary for the degrees of freedom parameter through the prior distribution.

The Student-t specification in (2.1) needs to be re-written in order to perform a convenient Bayesian estimation (see, e.g. Geweke, 1993):

$$\begin{aligned} y_t = {} & \varepsilon_t (\varpi_t \varrho\, \boldsymbol{e}_t' \boldsymbol{h}_t)^{1/2} && \text{for } t = 1, \ldots, T \\ \varepsilon_t & \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \quad ; && \varpi_t \overset{\text{i.i.d.}}{\sim} \mathcal{IG}\left(\tfrac{\nu}{2}, \tfrac{\nu}{2}\right)\,, \end{aligned}$$

where $\mathcal{N}(0,1)$ is the standard Normal and $\mathcal{IG}$ the Inverted Gamma density. The degrees of freedom parameter ν characterizes the density of ϖ_t as follows:

$$p(\varpi_t \mid \nu) = \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \varpi_t^{-\frac{\nu}{2}-1} \exp\left[-\frac{\nu}{2\varpi_t}\right]. \tag{2.3}$$

[2] A sensitivity analysis has been performed in which the maximum value of the threshold was set to a small positive value. Results were however similar to those obtained by setting the maximum value at zero.

[3] The assumption $\boldsymbol{h}_0 \doteq \mathbf{0}$ could be relaxed, but for a large number of observations which is often the case with financial data, this should have a negligible impact on the posterior distribution. A sensitivity analysis has been performed and did not show significant differences in the posterior results.

For a parsimonious expression of the likelihood function, we define the vectors $\boldsymbol{y} \doteq (y_1 \cdots y_T)'$, $\boldsymbol{\varpi} \doteq (\varpi_1 \cdots \varpi_T)'$, $\boldsymbol{s} \doteq (s_1 \cdots s_T)'$ and $\boldsymbol{\alpha} \doteq (\boldsymbol{\alpha}_0' \, \boldsymbol{\alpha}_1' \, \boldsymbol{\alpha}_2')'$. The model parameters are then regrouped into $\Theta \doteq (\psi, \boldsymbol{\varpi}, \boldsymbol{s})$, where $\psi \doteq (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \nu, P)$. Finally, we define the diagonal matrix $\Sigma \doteq \Sigma(\Theta) = \text{diag}\,(\{\varpi_t \varrho\, \boldsymbol{e}_t' \boldsymbol{h}^t\}_{t=1}^T)$ where we recall that $\varrho, \boldsymbol{e}_t$ and $\boldsymbol{h}_t$ are both functions of the model parameters. We can now express the likelihood function of Θ as follows:

$$\mathcal{L}(\Theta|\boldsymbol{y}) \propto (\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{y}'\Sigma^{-1}\boldsymbol{y}\right].$$

This likelihood function is invariant with respect to relabelling the states, which leads to a lack of identification of the state-specific parameters. So, without a prior inequality restriction on some state-specific parameters, a multimodal posterior is obtained and is difficult to interpret and summarize. To overcome this problem, we make use of the permutation sampler of Frühwirth-Schnatter (2001) to find suitable identification constraints. The permutation sampler requires priors that are labelling invariant. Furthermore, we cannot be completely non-informative about the state specific parameters since, from a theoretical viewpoint, this would result in improper posteriors (see, e.g. Diebolt and Robert, 1994). In addition, a diffuse prior on $\boldsymbol{\tau}$ would lead to a flat posterior since the threshold parameters are unidentified for $\boldsymbol{\alpha}_2 = \boldsymbol{0}$.

For the scedastic function's parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\tau}$, we use truncated Normal densities:

$$\begin{aligned}
p(\boldsymbol{\alpha}) &\propto \mathcal{N}_{3K}(\boldsymbol{\alpha}|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \Sigma_{\boldsymbol{\alpha}})\,\mathbb{I}\{\boldsymbol{\alpha} \geqslant \boldsymbol{0}\}\\
p(\boldsymbol{\beta}) &\propto \mathcal{N}_{K}(\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})\,\mathbb{I}\{\boldsymbol{\beta} \geqslant \boldsymbol{0}\}\\
p(\boldsymbol{\tau}) &\propto \mathcal{N}_{K}(\boldsymbol{\tau}|\boldsymbol{\mu}_{\boldsymbol{\tau}}, \Sigma_{\boldsymbol{\tau}})\,\mathbb{I}\{\boldsymbol{\tau}_{\min} \leqslant \boldsymbol{\tau} \leqslant \boldsymbol{0}\},
\end{aligned}$$

where $\boldsymbol{\mu}_\bullet$, $\Sigma_\bullet$ and $\boldsymbol{\tau}_{\min}$ are the hyperparameters and $\mathcal{N}_d$ is the d-dimensional Normal density ($d > 1$). The assumption of labelling invariance is fulfilled if we assume that the hyperparameters are the same for all states. In particular, we set $[\boldsymbol{\mu}_{\boldsymbol{\alpha}}]_i \doteq \mu_{\alpha_0}$, $[\Sigma_{\boldsymbol{\alpha}}]_{ii} \doteq \sigma^2_{\alpha_0}$, $[\boldsymbol{\mu}_{\boldsymbol{\beta}}]_i \doteq \mu_{\boldsymbol{\beta}}$, $[\Sigma_{\boldsymbol{\beta}}]_{ii} \doteq \sigma^2_{\boldsymbol{\beta}}$, $[\boldsymbol{\mu}_{\boldsymbol{\tau}}]_i \doteq \mu_{\boldsymbol{\tau}}$, $[\Sigma_{\boldsymbol{\tau}}]_{ii} \doteq \sigma^2_{\boldsymbol{\tau}}$, $[\boldsymbol{\tau}_{\min}]_i \doteq \tau_{\min}$ for $i = 1, \ldots, K$; $[\boldsymbol{\mu}_{\boldsymbol{\alpha}}]_i \doteq \mu_{\alpha_1}$, $[\Sigma_{\boldsymbol{\alpha}}]_{ii} \doteq \sigma^2_{\boldsymbol{\alpha}_1}$ for $i = K+1, \ldots, 2K$ and $[\boldsymbol{\mu}_{\boldsymbol{\alpha}}]_i \doteq \mu_{\alpha_2}$, $[\Sigma_{\boldsymbol{\alpha}}]_{ii} \doteq \sigma^2_{\boldsymbol{\alpha}_2}$ for $i = 2K+1, \ldots, 3K$, where $\mu_\bullet, \sigma^2_\bullet$ and $\tau_{\min}$ are fixed values. Note that a lower boundary $\tau_{\min}$ is used since the likelihood function is invariant for threshold values below the minimum value of the observed data. Hence, the prior on $\boldsymbol{\tau}$ could be considered to correspond to an empirical Bayes approach rather than a fully Bayesian one.

The prior density of the vector $\boldsymbol{\varpi}$ conditional on ν is found by noting that ϖ_t are independent and identically distributed from (2.3), which yields:

$$p(\boldsymbol{\varpi}|\nu) = \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \left(\prod_{t=1}^T \varpi_t\right)^{-\frac{\nu}{2}-1} \exp\left[-\frac{1}{2}\sum_{t=1}^T \frac{\nu}{\varpi_t}\right].$$

Following Deschamps (2006), we choose a translated Exponential with parameters $\lambda > 0$ and $\delta \geqslant 2$ for the prior on the degrees of freedom parameter:

$$p(\nu) = \lambda \exp[-\lambda(\nu - \delta)]\,\mathbb{I}\{\delta < \nu < \infty\}.$$

For large values of λ, the mass of the prior is concentrated in the neighbourhood of δ and a constraint on the degrees of freedom can be imposed in this manner. The Normality for the errors is obtained when δ becomes large. As pointed out by Deschamps (2006), this prior density is useful for two reasons. First, for numerical reasons, to bound the degrees of freedom parameter away from two which avoids explosion of the conditional variance. Second, we can approximate the Normality for the errors while maintaining a reasonably tight prior which can improve the convergence of the MCMC sampler.

Conditionally on the transition probabilities matrix P, the prior on vector $\boldsymbol{s}$ is Markov:

$$p(\boldsymbol{s}|P) = \pi(s_1) \prod_{i=1}^{K} \prod_{j=1}^{K} P_{ij}^{N_{ij}},$$

where $N_{ij} \doteq \#\{s_{t+1} = j | s_t = i\}$ is the number of one-step transitions from state i to j in the vector $\boldsymbol{s}$. The mass function for the initial state, $\pi(s_1)$, is obtained by calculating the ergodic probabilities of the Markov chain (see, e.g. Hamilton, 1994, section 22.2).

The prior density for the transition matrix is obtained by assuming that the K rows are independent and that the density of the ith row is Dirichlet with parameter $\boldsymbol{\eta}_i \doteq (\eta_{i1} \cdots \eta_{iK})$:

$$p(P) = \prod_{i=1}^{K} \mathcal{D}(\boldsymbol{\eta}_i) \propto \prod_{i=1}^{K} \prod_{j=1}^{K} P_{ij}^{\eta_{ij}-1}.$$

Due to the labelling invariance assumption, we require that $\eta_{ii} \doteq \eta_p$ for $i = 1, \ldots, K$ and $\eta_{ij} \doteq \eta_q$ for $i, j \in \{1, \ldots, K; i \neq j\}$.

Finally, we form the joint prior by assuming prior independence between $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, (\boldsymbol{\varpi}, \nu)$ and $(\boldsymbol{s}, P)$. The joint posterior density is then obtained by combining the likelihood function and the joint prior via Bayes' rule.

3. SIMULATING THE JOINT POSTERIOR

We draw an initial value from an arbitrary proper distribution and then we cycle through the full conditionals:

$$\begin{array}{lll} p(\boldsymbol{\alpha}|\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y}) & p(\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y}) & p(\boldsymbol{\tau}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y}) \\ p(\boldsymbol{\varpi}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \nu, \boldsymbol{s}, \boldsymbol{y}) & p(\nu|\boldsymbol{\varpi}) & \\ p(\boldsymbol{s}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, P, \boldsymbol{y}) & p(P|\boldsymbol{s}), & \end{array}$$

using the most recent conditional values. Among the full conditional densities listed above, only $\boldsymbol{\varpi}$ and P can be simulated from known expressions. Draws of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are achieved by a multivariate extension of the methodology proposed by Nakatsuma (2000). The generation of state vector $\boldsymbol{s}$ is made by using the FFBS algorithm described in Chib (1996). Finally, sampling ν is achieved by an efficient rejection technique. The reader is referred to the Appendix for further details on the simulation techniques.

As noted previously, we rely on the permutation sampler of Frühwirth-Schnatter (2001) to overcome the identification problem encountered with mixture models.[4] We use the random permutation sampler to determine suitable identification constraints; in this version of the permutation sampler, each pass of the MCMC scheme is followed by a random permutation of the regime definition. This algorithm improves the mixing of the MCMC sampler and allows

[4] The permutation-augmented sampler by Geweke (2007) can also be used to that purpose. The constrained permutation sampler is however useful as a diagnostic test to determine whether the constraint is well suited.

to explore the full unconstrained parameter space. Then, we post-process the MCMC output of the random permutation sampler in an exploratory way to determine appropriate identification constraint. At this stage, the model parameters are estimated again under the constraint by enforcing the corresponding permutation of the regimes; this version of the permutation sampler is referred to as the constrained permutation sampler.

4. AN APPLICATION TO THE SWISS MARKET INDEX

We apply our Bayesian estimation method to demeaned daily log-returns $\{y_t\}$ of the Swiss Market Index (SMI). The sample period is from November 12, 1990 to April 10, 2006 for a total of 4000 observations and the log-returns are expressed in percent. The first 2500 observations, which represent slightly less than two third of the sample, are used to estimate the models while the remaining 1500 log-returns are used in a forecasting performance analysis.

The time series under investigation is plotted in Figure 1. We test for autocorrelation in the times series by testing the joint nullity of autoregressive coefficients for $\{y_t\}$. We estimate the regression with autoregression coefficients up to lag 15 and compute the covariance matrix using the White estimate. The p-value of the Wald test is 0.1427, which does not support the presence of autocorrelation. When testing for the autocorrelation in the series of squared observations we strongly reject the null of the absence of autocorrelation. Thus, as is typical for financial data, the presence of GARCH effects in the time series of returns is confirmed.

4.1. Estimation

We apply our Bayesian estimation to a single-regime threshold GJR(1,1) model and a two-state Markov-switching threshold GJR(1,1) model, henceforth referred to as TGJR and MSTGJR. In order to diminish the impact of the prior on the joint posterior, we estimate both models under rather vague priors. For the priors on the scedastic function's parameters, we set $\mu_\bullet$ to zero, $\sigma^2_\bullet$ to 10,000 and $\tau_{\min}$ to -1.38, which is the 2.5th percentile of the observed log-returns. For the prior on the degrees of freedom parameter, we choose $\lambda = 0.01$ and $\delta = 2$ to ensure the existence of the conditional variance. Finally, we set $\eta_{ii} = 2$ and $\eta_{ij} = \eta_{ji} = 1$ for $i, j \in \{1, 2\}$ so that we have a prior belief that the probabilities of persistence are bigger than the probabilities of transition. A sensitivity analysis has been performed and confirmed that our initial choice is vague enough and does not introduce significant information in our estimation.

We run two chains for 50,000 iterations each and assess the convergence of the sampler by using the diagnostic test of Gelman and Rubin (1992). The convergence appears rather quickly, but we nevertheless consider the first half of the iterations as a burn-in phase for precaution. For the TGJR model, the acceptance rate is 88% for $\boldsymbol{\alpha}$ and 97% for β, indicating that the proposal distributions for these parameters are close to the full conditionals. The acceptance rate for τ is 32%. The one-lag autocorrelations in the chain range from 0.34 for α_1 to 0.94 for β which is reasonable. For the MSTGJR model, the random permutation sampler is run first to determine suitable identification constraints. In Figure 2, we show the contour plots of the posterior density for (β^k, α_0^k), (β^k, α_1^k), (β^k, α_2^k) and (β^k, τ^k), respectively.[5] As we can notice, the bimodality of the posterior density is clear for the parameter β^k on the four graphs, suggesting a constraint of the type $\beta^1 < \beta^2$ for identification. Therefore, the model is estimated again under this constraint;

[5] The value k is arbitrary since all marginal distributions contain the same information (Frühwirth-Schnatter, 2001).

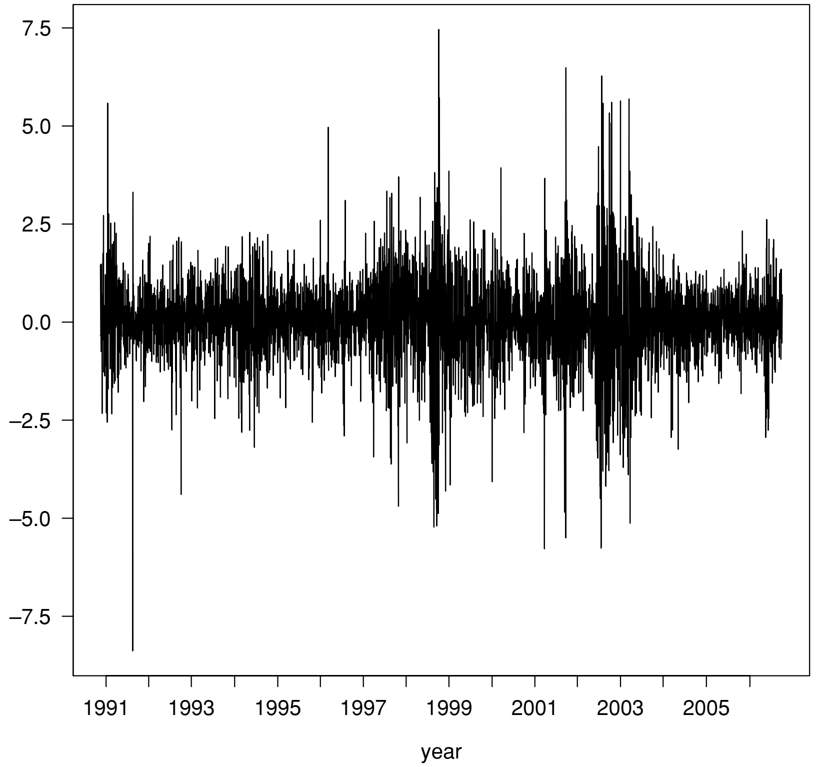

Figure 1. Swiss Market Index daily log-returns (in %).

in that case, label switching only appeared 26 times after the burn-in phase thus confirming the suitability of the identification constraint. The acceptance rates obtained with the constrained permutation sampler range from 15% for $\boldsymbol{\tau}$ to 93% for $\boldsymbol{\beta}$. The one-lag autocorrelations range from 0.7 for α_1^2 to 0.97 for α_0^1. We keep every fifth draw from the MCMC output for both models in order to diminish the autocorrelation in the chains. The two chains are then merged to get a final sample of length 10,000. Our algorithm is not only fully automatic, the results described here also show desirable convergence properties.

Note that a three-state MSTGJR model has also been estimated. However post-processing the MCMC output has not allowed to find any clear identification constraint, which suggests that the number of regime is too large (Frühwirth-Schnatter, 2006, section 4.2). Moreover, the model likelihood estimate clearly indicates that a three-state Markov-switching model is not supported by the data (see Table 3).

The posterior statistics for both models are reported in Table 1. In the case of the TGJR model (upper panel), we note the presence of an asymmetric response to past shocks; the posterior mean of α_2 is 0.174 and the probability $\mathbb{P}(\alpha_2 > 0 | \boldsymbol{y})$ is one. The posterior mean of β is 0.803, indicating a high memory in the conditional variance process. The posterior mean and median of the threshold parameter are -0.094 and -0.071, respectively. However, the 95% highest posterior density interval (HPDI) for τ does contain the value of zero, suggesting that the threshold value

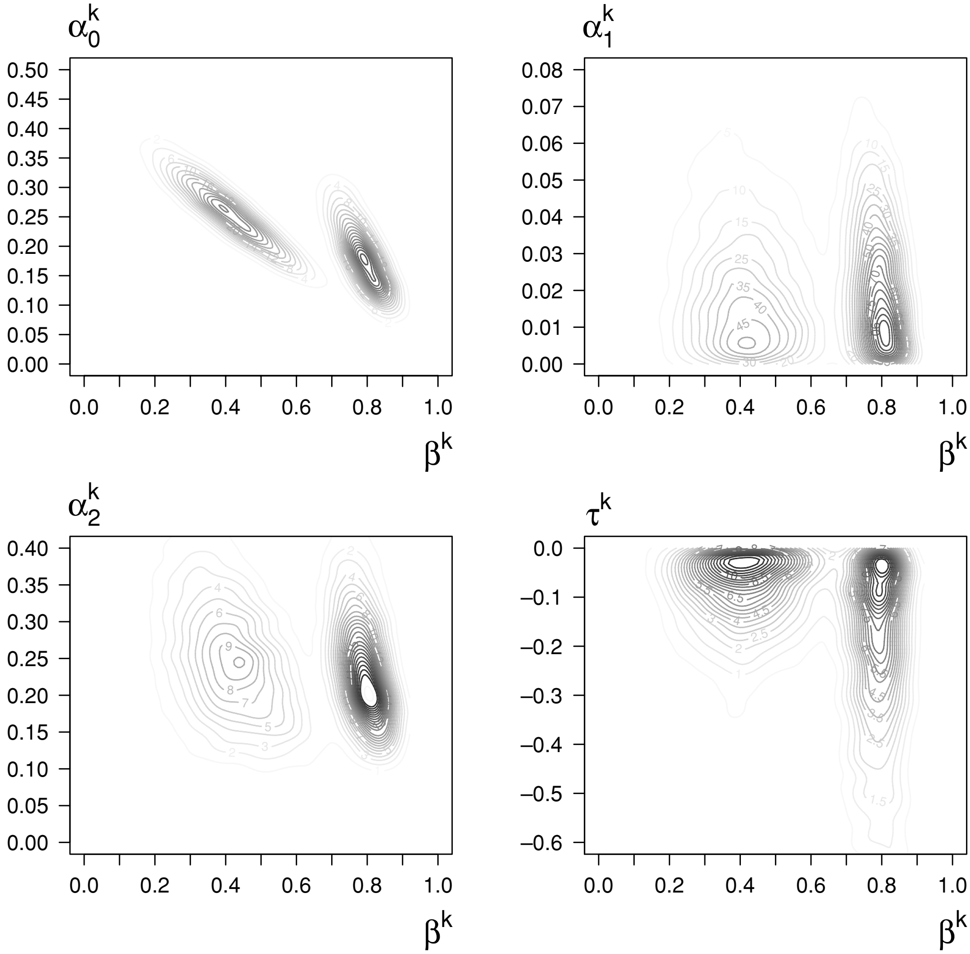

Figure 2. Contour plots for (β^k, α_0^k), (β^k, α_1^k), (β^k, α_2^k) and (β^k, τ^k).

is not significantly different from zero at the 5% level. Finally, the value of 8.64 for the posterior mean of the degrees of freedom parameter indicates conditional leptokurtosis in the data set.

In the MSTGJR case (lower panel), we observe an asymmetric response to past shocks on the conditional variance for both states; the posterior mean of α_2^k is 0.253 and 0.243 for the first and the second state, respectively; both estimated probabilities $\mathbb{P}(\alpha_2^k > 0 | \boldsymbol{y})$ are equal to one. A comparison of the scedastic function's parameters between the two regimes indicates similar 95% HPDI for the components of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ while the difference for components of $\boldsymbol{\alpha}_0$ is slightly more pronounced. The 95% HPDI for parameters α_1^1 and α_1^2 does contain the value of zero. Moreover, the 95% HPDI of $(\alpha_2^1 - \alpha_2^2)$ (not shown) does contain the value of zero, suggesting that the asymmetric function is similar for both regimes. We note that the difference between the two regimes is significant for β since the 95% HPDI intervals do not overlap. The posterior mean of β^1 is 0.411, indicating a low memory in the variance process, while the posterior mean of β^2 is 0.788, similar to the value observed in the single-regime model. For the threshold parameter, the

Table 1. Estimation results.

ψ	$\overline{\psi}$	$\psi_{0.5}$	[95% HPDI]		NSE	IF
TGJR						
α_1	0.058	0.057	0.024	0.095	0.242	1.74
α_2	0.174	0.170	0.100	0.257	0.598	2.18
τ	−0.094	−0.071	−0.267	0.000	2.179	6.81
β	0.803	0.804	0.745	0.856	1.118	15.25
ν	8.648	8.502	6.348	11.100	42.178	11.19
MSTGJR						
α_0^1	0.256	0.256	0.151	0.358	2.946	29.39
α_0^2	0.185	0.179	0.088	0.295	1.984	12.24
α_1^1	0.023	0.017	0.000	0.063	0.325	2.76
α_1^2	0.024	0.020	0.000	0.058	0.290	2.45
α_2^1	0.253	0.245	0.096	0.421	2.007	5.27
α_2^2	0.243	0.229	0.109	0.394	1.957	5.82
τ^1	−0.089	−0.062	−0.247	0.000	3.214	11.71
τ^2	−0.223	−0.172	−0.616	0.000	11.664	35.34
β^1	0.411	0.410	0.208	0.627	4.126	14.41
β^2	0.788	0.791	0.700	0.872	1.882	17.02
ν	10.020	9.866	7.097	12.930	54.942	12.61
p_{11}	0.996	0.997	0.993	1.000	0.023	1.22
p_{22}	0.995	0.996	0.990	0.999	0.027	1.17

Note: $\overline{\psi}$: posterior mean; $\psi_{0.5}$: posterior median; [95% HPDI]: 95% highest posterior density interval; NSE: numerical standard error ($\times 10^3$) and IF: inefficiency factor (i.e. ratio of the squared numerical standard error and the variance of the sample mean from a hypothetical i.i.d. sampler). The posterior statistics are based on 10,000 draws from the constrained posterior sample.

posterior mean is −0.089 in the first regime and −0.223 in the second regime. However, the 95% HPDI for $(\tau^1 - \tau^2)$ (not shown) does contain the value of zero, indicating that the thresholds are not significantly different at the 5% level. Overall, these results suggest that both the asymmetric responses and the locations of the asymmetry are similar between the two regimes.[6] As in the single-regime model, the posterior distribution for the degrees of freedom parameter indicates conditional leptokurtosis. We note however that the posterior mean and median are larger than for the TGJR model. The posterior means for probabilities p_{11} and p_{22} are, respectively 0.996 and 0.995, indicating infrequent mixing between states. The 95% HPDI of $(p_{11} - p_{22})$ (not shown) does contain the value of zero. Finally, the inefficiency factors (IF) reported in the last column of Table 1 indicate that using 10,000 draws out of the MCMC sampler seems appropriate

[6] We could continue with a reduced model, i.e. by setting $\alpha_1^1 = \alpha_1^2, \alpha_2^1 = \alpha_2^2$ and $\tau^1 = \tau^2$. However, this is not the goal of the paper to perform such model selection.

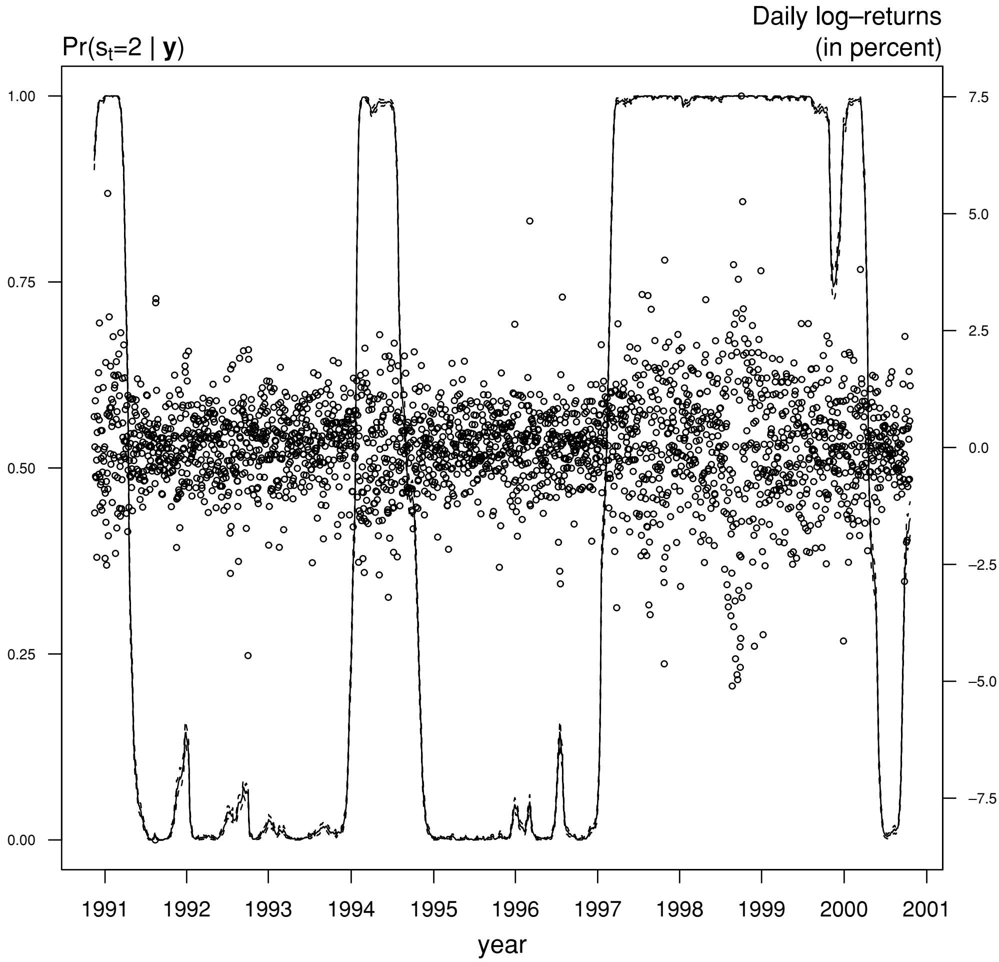

Figure 3. Smoothed probabilities for the second regime (solid line, left-hand axis) together with the in-sample log-returns (circles, right-hand axis).

if we require that the Monte Carlo error in estimating the mean is smaller than one percent of the variation of the error due to the data.[7]

In Figure 3, we present the smoothed probabilities for the second regime (solid line, left-hand axis) together with the in-sample daily log-returns (circles, right-hand axis). The 95% (robust) confidence bands are shown in dashed lines but are almost indistinguishable from the point estimates. As we can notice, the second state is clearly associated with high-volatility periods, while the first regime corresponds to more tranquil periods. The beginning of the year 1991 is

[7] The inefficiency factors are computed as the ratio of the squared numerical standard error (NSE) of the MCMC simulations and the variance estimate divided by the number of iterations. The NSE are estimated by the method of Andrews (1991), using a Parzen kernel and AR(1) pre-whitening as presented in Andrews and Monahan (1992). Note also that the inefficiency factor is the inverse of the relative numerical efficiency (RNE) introduced by Geweke (1989).

associated with the high-volatility state. Then, from the second half of 1991 to 1997, the returns are clearly associated with the low-volatility regime, with the exception of 1994. From 1997 to 2000, the model remains in the high-volatility regime with a transition during the second semester 2000 to the low-volatility state.

4.2. *In-sample performance analysis*

4.2.1. Model diagnostics. We check for model misspecification by analyzing the predictive probabilities also referred to as p-scores (see, e.g. Kaufmann and Frühwirth-Schnatter, 2002). We make use of a simpler version of this method, as proposed by Kim et al. (1998), which consists in conditioning on point estimates of $\psi \doteq (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \nu, P)$. To be meaningful, the point estimate has to be chosen when the identification is imposed. Hence, we consider the mean $\overline{\psi}$ of the constrained posterior sample. Upon defining $\mathcal{F}_{t-1}$ as the information set up to time $(t-1)$, the (approximate) p-scores are defined as follows:

$$z_t \doteq \sum_{k=1}^{K} \mathbb{P}(Y_t \leqslant y_t | s_t = k, \overline{\psi}, \mathcal{F}_{t-1}) \, \mathbb{P}(s_t = k | \overline{\psi}, \mathcal{F}_{t-1}) \, .$$

The probability $\mathbb{P}(Y_t \leqslant y_t | s_t = k, \overline{\psi}, \mathcal{F}_{t-1})$ can be estimated by the Student-t integral and the filtered probability $\mathbb{P}(s_t = k | \overline{\psi}, \mathcal{F}_{t-1})$ is obtained as a byproduct from the FFBS algorithm (see, e.g. Chib, 1996, p. 83). Under a correct specification, the p-scores should have independent uniform distributions asymptotically (Rosenblatt, 1952). A further transformation through the Normal integral is often applied for convenience, i.e. $u_t \doteq \Phi^{-1}(z_t)$ where $\Phi^{-1}(\bullet)$ denotes the inverse cumulative standard Normal function. If the model is correct, $\{u_t\}$ should be independent standard Normal and common tests can be used to check these features. In particular, we test the presence of autocorrelation in the series $\{u_t\}$ and $\{u_t^2\}$ using a Wald test. We also report the results of a joint test for zero mean, unit variance, zero skewness, and the absence of excess kurtosis (Berkowitz, 2001).

In the case of the TGJR model, the Wald statistic for testing the joint nullity of autoregressive coefficients up to lag 15 for u_t has a p-value of 0.0903, and for u_t^2, a p-value of 0.3838. In the case of the MSTGJR model, the p-values are 0.0841 and 0.3909, respectively. Therefore, both models adequately capture the volatility clustering present in the data. The Normality test of Berkowitz (2001) yields p-values of 0.1184 for the TGJR model and 0.0756 for the MSTGJR model. Overall, these results indicate no evidence of misspecification at the 5% level for both models.

4.2.2. Deviance information criterion. In order to evaluate the goodness-of-fit of the models, we first use the Deviance information criterion (DIC) introduced by Spiegelhalter et al. (2002). Given a set of models, the one with the smallest DIC has the best balance between goodness-of-fit and model complexity.

As noted in Celeux et al. (2006), difficulties arise when applying this criterion to mixture models. To overcome these problems, we integrate out the state vector by considering the observed likelihood and make use of the constrained posterior sample in the estimation (Celeux et al., 2006, section 3.1). In the context of Markov-switching models, the observed likelihood is

Table 2. Deviance information criterion.

Model	DIC	95% CI
TGJR	6762.8	[6762.3 6763.3]
MSTGJR	6706.6	[6705.8 6707.1]

Note: 95% CI: 95% confidence interval of the DIC computed by a resampling technique.

(see, e.g. Kaufmann and Frühwirth-Schnatter, 2002, p. 457):

$$\mathcal{L}(\psi \,|\, \boldsymbol{y}) = \prod_{t=1}^{T} \left[\sum_{k=1}^{K} p(y_t | \psi, s_t = k, \mathcal{F}_{t-1}) \, \mathbb{P}(s_t = k | \psi, \mathcal{F}_{t-1}) \right], \tag{4.1}$$

and DIC $\doteq 2 \{\ln \mathcal{L}(\overline{\psi} \,|\, \boldsymbol{y}) - 2\, \mathbb{E}_{\psi|\boldsymbol{y}}[\ln \mathcal{L}(\psi \,|\, \boldsymbol{y})]\}$.

The DIC estimates are reported in Table 2 together with their 95% confidence intervals obtained by a resampling technique (Ardia, 2008, section 7.4.2). From this table, we can notice that the criterion favours the MSTGJR model. Indeed, the DIC estimates based on the initial joint posterior sample is 6762.8 for the TGJR model and 6706.6 for the MSTGJR model. Both 95% confidence intervals do not overlap which suggests significant improvement of the Markov-switching model.

4.2.3. Model likelihood. As a second criterion to discriminate between the models, we consider the model likelihood:

$$p(\boldsymbol{y}) = \int \mathcal{L}(\psi \,|\, \boldsymbol{y}) p(\psi) d\psi \,,$$

where $\mathcal{L}(\psi \,|\, \boldsymbol{y})$ is given in (4.1) and $p(\psi)$ is the joint prior density on ψ. The estimation of $p(\boldsymbol{y})$ requires the integration over the whole set of parameters, which is a difficult task in practice. In the context of mixture models, Frühwirth-Schnatter (2004) documents that the bridge sampling technique (Meng and Wong, 1996) using the MCMC output of the random permutation sampler and an i.i.d. sample from an importance density $q(\psi)$ which approximates the unconstrained posterior yields the best (and a robust) estimator. Specifically, the model likelihood is approximated as the limit of:

$$p_t(\boldsymbol{y}) \doteq p_{t-1}(\boldsymbol{y}) \times \frac{\frac{1}{L} \sum_{l=1}^{L} \frac{p_{t-1}(\psi^{[l]} | \boldsymbol{y})}{Lq(\psi^{[l]}) + Mp_{t-1}(\psi^{[l]} \,|\, \boldsymbol{y})}}{\frac{1}{M} \sum_{m=1}^{M} \frac{q(\psi^{[m]})}{Lq(\psi^{[m]}) + Mp_{t-1}(\psi^{[m]} | \boldsymbol{y})}} \,,$$

where $\{\psi^{[m]}\}_{m=1}^{M}$ are MCMC draws from the joint posterior, $\{\psi^{[l]}\}_{l=1}^{L}$ are i.i.d. draws from the importance sampling density $q(\psi)$ and $p_t(\psi \,|\, \boldsymbol{y}) \doteq \mathcal{L}(\psi \,|\, \boldsymbol{y}) p(\psi) / p_t(\boldsymbol{y})$. This sequence is typically initialized with the reciprocal importance sampling estimator of Gelfand and Dey (1994).

We adopt the approach of Kaufmann and Frühwirth-Schnatter, (2002, pp. 438–439) to construct an importance density which reproduce the $K!$ modes of the unconstrained posterior. Specifically, $q(\psi)$ is obtained from the MCMC output of the random permutation sampler using

Table 3. Model likelihood estimators.

Model	$\ln p_0(\boldsymbol{y})$	$\ln p(\boldsymbol{y})$
TGJR	−3405.33 (2.92)	−3408.04 (2.85)
MSTGJR	−3394.04 (3.19)	−3401.00 (3.40)
MSTGJR (three states)	−3427.82 (8.73)	−3492.74 (9.89)

Note: $\ln p_0(\boldsymbol{y})$: natural logarithm of the model likelihood estimate using reciprocal importance sampling; $\ln p(\boldsymbol{y})$: natural logarithm of model likelihood estimate using bridge sampling; ($\bullet$) numerical standard error of the estimator ($\times 10^2$).

a mixture of the proposal and conditional densities:

$$\begin{aligned} q(\psi) \doteq \Bigg[\frac{1}{R} \sum_{r=1}^{R} & q_{\boldsymbol{\alpha}} \left(\boldsymbol{\alpha} | \boldsymbol{\alpha}^{[r]}, \boldsymbol{\beta}^{[r]}, \boldsymbol{\tau}^{[r]}, \boldsymbol{\varpi}^{[r]}, \nu^{[r]}, \boldsymbol{s}^{[r]}, \boldsymbol{y} \right) \\ & \times q_{\boldsymbol{\beta}} \left(\boldsymbol{\beta} | \boldsymbol{\alpha}^{[r]}, \boldsymbol{\beta}^{[r]}, \boldsymbol{\tau}^{[r]}, \boldsymbol{\varpi}^{[r]}, \nu^{[r]}, \boldsymbol{s}^{[r]}, \boldsymbol{y} \right) \\ & \times q_{\boldsymbol{\tau}} \left(\boldsymbol{\tau} | \boldsymbol{\alpha}^{[r]}, \boldsymbol{\beta}^{[r]}, \boldsymbol{\tau}^{[r]}, \boldsymbol{\varpi}^{[r]}, \nu^{[r]}, \boldsymbol{s}^{[r]}, \boldsymbol{y} \right) \times p(P | \boldsymbol{s}^{[r]}) \Bigg] \times q_{\nu}(\nu) \,, \end{aligned}$$

where $q_{\boldsymbol{\alpha}}, q_{\boldsymbol{\beta}}$ and $q_{\boldsymbol{\tau}}$ are the proposal densities for $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\tau}$, respectively, and $p(P|\boldsymbol{s}^{[r]})$ is the product of Dirichlet posterior densities for the transition probabilities (see the Appendix). q_{ν} is a truncated skewed Student-t density whose parameters are estimated by ML from the posterior sample (Ardia, 2008, section 7.4.3).

In Table 3, we report the natural logarithm of the model likelihoods obtained using the reciprocal importance sampling estimator (second column) and the bridge sampling estimator (last column) for $M = L = 1000$ draws with $R = 1000$ components. From this table, we can notice that both estimators are higher for the MSTGJR model, indicating a better in-sample fit for the regime-switching specification. As an additional discrimination criterion, we compute the (transformed) Bayes factor in favour of the MSTGJR model (see, e.g. Kass and Raftery, 1995, section 3.2). The estimated value is $2 \times \ln \mathrm{BF} = 2 \times [-3401.00 - (-3408.04)] = 10.08$, which strongly supports the in-sample evidence in favour of the regime-switching model.[8]

4.3. *Forecasting performance analysis*

In order to evaluate the ability of the competing models to predict the future behaviour of the volatility process, we first study the forecasted one-day ahead Value at Risk (VaR), which is a common tool to measure financial and market risks. The one-day ahead VaR at risk level $\phi \in (0, 1)$, VaR^{ϕ}, is estimated by calculating the $(1 - \phi)$th percentile of the one-day ahead predictive

[8] The model likelihood is sensitive to the choice of the prior distribution, so we must test whether an alternative joint prior specification would have modified the conclusion of our analysis. To answer this question, we modified the hyperparameters' value and ran the sampler again. We considered slightly more informative priors for the scedastic function's parameters. As an alternative prior on the degrees of freedom parameter we chose $\lambda = 0.02$ and $\delta = 2$ which implies a prior mean of 52. Finally, the hyperparameters for the prior on the transition probabilities were set to $\eta_{ii} = 3$ and $\eta_{ij} = \eta_{ji} = 1$ for $i, j \in \{1, 2\}$. The results were similar to those obtained previously, confirming the better fit of the MSTGJR model.

distribution which is obtained by simulation from the joint posterior sample.[9] For the two models, the predictions are obtained for the out-of-sample window which consists of 1500 daily log-returns. To verify the accuracy of the VaR estimates, we adopt the testing methodology of Christoffersen (1998). This approach is based on the study of the random sequence $\{V_t^\phi\}$ where $V_t^\phi \doteq \mathbb{I}\{y_{t+1<\mathrm{VaR}_t^\phi}\}$ if $\phi > 0.5$ and $V_t^\phi \doteq \mathbb{I}\{y_{t+1>\mathrm{VaR}_t^\phi}\}$ if $\phi \leqslant 0.5$. A sequence of VaR forecasts at risk level ϕ has correct conditional coverage if $\{V_t^\phi\}$ is an independent and identically distributed sequence of Bernoulli random variables with parameter $(1-\phi)$ if $\phi > 0.5$ and with parameter ϕ if $\phi \leqslant 0.5$. This hypothesis can be verified by testing jointly the independence on the series and the unconditional coverage of the VaR forecasts.[10]

The forecasting results of the VaR with typical (short and long positions) risk levels used in financial risk management are reported in Table 4. The second and third columns give the expected and observed number of violations. The last three columns report the p-values for the tests of correct unconditional coverage (UC), independence (IND) and correct conditional coverage (CC). From this table, we note that the observed number of violations for the MSTGJR model are closer to the expected values than for the TGJR model. Indeed, at the 5% significance level, the UC test is rejected two times for the MSTGJR model while it is rejected five times for the TGJR model. The IND test is rejected two times for both models. We can notice that for some risk levels this test is not applicable since no consecutive violations have been observed. The joint hypothesis of correct unconditional coverage and independent sequence is obtained via the CC test. In the case of the MSTGJR model, the CC test is (slightly) rejected for risk levels 0.9 and 0.075. The test is strongly rejected for the TGJR model at all risk levels except for $\phi = 0.1$.

As a second measure of the forecasting performance, we compute the natural logarithm of the predictive likelihood $p(y_{T+1} \dots y_{T+h} \mid \boldsymbol{y})$ for the two models, where $T = 2500$ and $h = 1500$. The estimation is obtained by Monte Carlo integration from the joint posterior sample (see, e.g. Geweke, 2005, section 2.6.2). The estimate is -5242.8 for the single-regime model and -5217.8 for the Markov-switching model. The numerical standard errors of the estimates, obtained by a resampling technique (Ardia, 2008, section 7.4.2), are 0.152 and 0.173, respectively. Hence, the predictive likelihood for the MSTGJR model is significantly higher than for the TGJR model. Overall, these results indicate the better out-of-sample performance of the Markov-switching model compared to the single-regime specification.

5. CONCLUSION

MSGARCH models provide an explanation for the high persistence in volatility observed in single-regime GARCH models and allow for a sudden change in the (unconditional)

[9] In order to simulate from the predictive distribution over the out-of-sample observations window, the posterior sample of Θ should be updated using the most recent information. As a consequence, forecasting the one-day ahead VaR would necessitate the estimation of the joint posterior sample at each time point in the out-of-sample observation window. However, such an approach is computationally impractical for large data set such as ours. Combination of MCMC and importance sampling to estimate efficiently this predictive distribution is proposed by Gerlach et al. (1999). Nevertheless, for the sake of simplicity, we consider the same joint posterior sample, based on the in-sample data set, when forecasting the VaR.

[10] A (conservative) joint test for multiple VaR risk levels can be obtained using the Bonferroni correction. Specifically, a joint test of k VaR risk levels at significance level s is rejected if the smallest p-value among the k individual tests is smaller than s/k.

Table 4. Results of the VaR.

ϕ	$\mathbb{E}(V_t^\phi)$	#	UC	IND	CC
TGJR					
0.99	15.0	17	0.611	NA	NA
0.975	37.5	41	0.568	NA	NA
0.95	75.0	97	0.013	0.124	0.013
0.925	112.5	130	0.093	0.024	0.019
0.9	150.0	161	0.349	0.002	0.006
0.1	150.0	129	0.065	0.288	0.103
0.075	112.5	88	0.013	0.276	0.025
0.05	75.0	50	0.002	NA	NA
0.025	37.5	13	0.000	NA	NA
0.01	15.0	5	0.003	NA	NA
MSTGJR					
0.99	15.0	15	1.000	NA	NA
0.975	37.5	36	0.803	NA	NA
0.95	75.0	87	0.165	0.105	0.103
0.925	112.5	115	0.807	0.134	0.315
0.9	150.0	149	0.931	0.014	0.048
0.1	150.0	149	0.931	0.075	0.203
0.075	112.5	101	0.252	0.022	0.038
0.05	75.0	62	0.113	NA	NA
0.025	37.5	18	0.000	NA	NA
0.01	15.0	5	0.003	NA	NA

Note: ϕ: risk level; $\mathbb{E}(V_t^\phi)$: expected number of violations; #: observed number of violations; UC: p-value for the correct unconditional coverage test; IND: p-value for the independence test; CC: p-value for the correct conditional coverage test and NA: not applicable.

volatility level which improves significantly the volatility forecasts. Different parametrizations of MSGARCH models have been proposed in the literature but they raise difficulties both in their estimation and their interpretation. Haas et al. (2004) overcome these problems by providing a model which can be estimated by ML and which allows for a clear-cut interpretation of the variance dynamics in each regime. This paper generalizes the model of Haas et al. (2004) in three ways. First, we account for asymmetric movements between the conditional variances and the underlying time series by replacing the separate GARCH(1,1) processes by GJR(1,1) processes. Second, a free non-positive threshold parameter is used to test whether it is the sign of the past shock or the effect of a particularly bad news which influences the conditional variance. Finally, we consider Student-t innovations instead of Normal innovations to gain flexibility in the modelling. The model is estimated under the Bayesian approach using a novel, fully automatic MCMC procedure. As an application, we fit a single-regime threshold GJR(1,1) model and a two-state Markov-switching threshold GJR(1,1) model to the Swiss Market Index log-returns. We show the presence of two distinct volatility regimes in the time series. Moreover, the posterior results indicate no difference between asymmetries and locations of the asymmetry for highly volatile periods relative to more subdued ones. Finally, we test the in- and out-of-sample performance of the two competing models and document the better performance of the Markov-switching specification.

In this study, we have considered a fixed transition matrix for the state process. As a consequence, the expected persistence of the regimes is constant over time, which is questionable. In a more general formulation, we could allow the transition probabilities to change over time depending on some observables (see, e.g. Gray, 1996). Also, the assumption of recurring states could be relaxed using the hierarchical approach of Pesaran et al. (2006). Finally, we could allow the degrees of freedom parameter to be state dependent as in Dueker (1997) and Perez-Quiros and Timmermann (2001). We leave these extensions for future research.

ACKNOWLEDGMENTS

This version has been written while the author was visiting the Econometric Institute, Erasmus University Rotterdam, the Netherlands. The author sincerely acknowledges the hospitality of Herman K. van Dijk and is grateful to the Swiss National Science Foundation (under grant #FN PB FR1-121441) for financial support. The author wishes to thank Luc Bauwens, Cathy W. Chen, Carlos Ordás Criado, Philippe J. Deschamps, Dennis Fok, Richard H. Gerlach, Lennart F. Hoogerheide, Richard Paap, Denis Pelletier, Jérôme Ph. Taillard, Herman K. van Dijk and Martin Wallmeier for helpful comments. He also acknowledges two anonymous reviewers and the Editor, Olivier Linton, for numerous helpful suggestions for improvement of the paper. Finally, the author thanks participants of the Econometric Institute seminars, Erasmus University Rotterdam, participants of the 2nd International Workshop on Computational and Financial Econometrics, University of Neuchâtel, and participants of the 14th International Conference on Computing in Economics and Finance, University la Sorbonne Paris. Any remaining errors or shortcomings are the author's responsibility.

REFERENCES

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–58.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*, 953–66.

Ardia, D. (2008). *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications*, *Volume 612*: *Lecture Notes in Economics and Mathematical Systems*. Heidelberg, Germany: Springer.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics 19*, 465–74.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics 3*, 167–79.

Celeux, G., F. Forbes, C. P. Robert and M. Titterington (2006). Deviance information criterion for missing data models. *Bayesian Analysis 1*, 651–706.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics 75*, 79–97.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*, 841–62.

Deschamps, P. J. (2006). A flexible prior distribution for Markov switching autoregressions with Student-t errors. *Journal of Econometrics 133*, 153–90.

Diebolt, J. and C. P. Robert (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society,* Series B *56*, 363–75.

Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business and Economic Statistics 15*, 26–34.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance 48*, 1749–78.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*, 194–209.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal 7*, 143–67.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. New York/Berlin/Heidelberg: Springer.

Gelfand, A. E. and D. K. Dey (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society,* Series B *56*, 501–14.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457–72.

Gerlach, R. H., C. Carter and R. Kohn (1999). Diagnostics for time series analysis. *Journal of Time Series Analysis 20*, 309–330.

Geweke, J. F. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–39.

Geweke, J. F. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics 8*, S19–40.

Geweke, J. F. (2005). *Contemporaneous Bayesian Econometrics and Statistics*. Series in Probability and Statistics. Hoboken, New Jersey, USA: John Wiley & Sons.

Geweke, J. F. (2007). Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics and Data Analysis 51*, 3529–50.

Geweke, J. F. and N. Terui (1991). Threshold autoregressive models for macroeconomic time series: A Bayesian approach. In *American Statistical Association 1991 Proceedings of the Business and Economic Statistics Section*, 42–50.

Glosten, L. R., R. Jaganathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 1779–801.

Gray, S. F. (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. S. Paolella (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics 2*, 493–530.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, USA: Princeton University Press.

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics 64*, 307–33.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*, 773–95.

Kaufmann, S. and S. Frühwirth-Schnatter (2002). Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis 23*, 425–58.

Kim, S., N. Shephard and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–93.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics 27*, 363–94.

Lamoureux, C. G. and W. D. Lastrapes (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economic Statistics 8*, 225–43.

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

Meng, X.-L. and W. H. Wong (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica 6*, 831–60.

Nakatsuma, T. (2000). Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach. *Journal of Econometrics 95*, 57–69.

Perez-Quiros, G. and A. Timmermann (2001). Business cycle asymmetries in stock returns: evidence from higher order moments and conditional densities. *Journal of Econometrics 103*, 259–306.

Pesaran, M. H., D. Pettenuzzo and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies 73*, 1057–84.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*, 470–72.

Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies 62*, 639–61.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society,* Series B *64*, 583–639.

APPENDIX: SIMULATING THE JOINT POSTERIOR

A.1. Generating $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$

The methodology used to draw $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\tau}$ can be viewed as a multivariate extension of the approach proposed by Nakatsuma (2000). Let us consider the vector $\boldsymbol{w}_t \doteq y_t^2 \boldsymbol{\iota}_K / \eta_t - \boldsymbol{h}_t$ where we define $\eta_t \doteq \varpi_t \varrho$ and $\boldsymbol{\iota}_K$ is a $K \times 1$ vector of ones. In order to simplify the notations further, we define $v_t \doteq y_t^2 / \eta_t$ which yields $\boldsymbol{w}_t = v_t \boldsymbol{\iota}_K - \boldsymbol{h}_t$. From there, we can transform the expression (2.2) as follows:

$$
\begin{aligned}
\boldsymbol{h}_t &= \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 y_{t-1}^2 + \boldsymbol{\alpha}_2 \odot \mathbf{I}\{y_{t-1} < \boldsymbol{\tau}\} \odot (\boldsymbol{\tau} - y_{t-1})^2 + \boldsymbol{\beta} \odot \boldsymbol{h}_{t-1} \\
\Leftrightarrow (v_t \boldsymbol{\iota}_K - \boldsymbol{w}_t) &= \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 y_{t-1}^2 + \boldsymbol{\alpha}_2 \odot \mathbf{I}\{y_{t-1} < \boldsymbol{\tau}\} \odot (\boldsymbol{\tau} - y_{t-1})^2 \\
&\quad + \boldsymbol{\beta} \odot (v_{t-1} \boldsymbol{\iota}_K - \boldsymbol{w}_{t-1}) \\
\Leftrightarrow \boldsymbol{w}_t &= v_t \boldsymbol{\iota}_K - \boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_1 y_{t-1}^2 - \boldsymbol{\alpha}_2 \odot \mathbf{I}\{y_{t-1} < \boldsymbol{\tau}\} \odot (\boldsymbol{\tau} - y_{t-1})^2 \\
&\quad - \boldsymbol{\beta} \odot v_{t-1} \boldsymbol{\iota}_K + \boldsymbol{\beta} \odot \boldsymbol{w}_{t-1} \,.
\end{aligned}
$$

Moreover, let us define $w_t \doteq \boldsymbol{e}_t' \boldsymbol{w}_t, h_t \doteq \boldsymbol{e}_t' \boldsymbol{h}_t$ and note that w_t can be written as follows:

$$
w_t \doteq \boldsymbol{e}_t' \boldsymbol{w}_t = v_t - h_t = \left(\frac{y_t^2}{\varpi_t \varrho h_t} - 1 \right) h_t = \left(\chi_1^2 - 1 \right) h_t \,,
$$

where χ_1^2 denotes a Chi-squared variable with one degree of freedom. This comes from the fact that the conditional distribution of y_t is Normal with zero mean and variance $\varpi_t \varrho h_t$. Therefore, the conditional mean of w_t is zero and the conditional variance is $2h_t^2$. As this is done for the single-regime GARCH model (Nakatsuma, 2000), this variable can be approximated by z_t, a Normal variable with a mean of zero and a variance of $2h_t^2$. The variable z_t can be further expressed as $z_t \doteq \boldsymbol{e}_t' \boldsymbol{z}_t$ where $\boldsymbol{z}_t$ is a function of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\tau}$

given by:

$$\begin{aligned} \boldsymbol{z}_t(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}) = v_t \boldsymbol{\iota}_K - \boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_1 y_{t-1}^2 - \boldsymbol{\alpha}_2 \odot \mathbf{I}\{y_{t-1} < \boldsymbol{\tau}\} \odot (\boldsymbol{\tau} - y_{t-1})^2 \\ - \boldsymbol{\beta} \odot v_{t-1} \boldsymbol{\iota}_K + \boldsymbol{\beta} \odot \boldsymbol{z}_{t-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}) . \end{aligned} \tag{A.1}$$

From there, we construct the vector $\boldsymbol{z} \doteq (z_1 \cdots z_T)'$ where $z_t \doteq \boldsymbol{e}_t' \boldsymbol{z}_t$ as well as the diagonal matrix $\Lambda \doteq \Lambda(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}) = \mathrm{diag}\Big(\{2\boldsymbol{e}_t' \boldsymbol{h}_t^2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau})\}_{t=1}^T\Big)$ and express the approximate likelihood function of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau})$ as follows:

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau} | \varpi, \nu, \boldsymbol{s}, \boldsymbol{y}) \propto (\det \Lambda)^{-1/2} \exp\left[-\frac{1}{2} \boldsymbol{z}' \Lambda^{-1} \boldsymbol{z}\right] . \tag{A.2}$$

The construction of the proposal densities for $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\tau}$ is based on this likelihood function.

A.1.1. Generating $\boldsymbol{\alpha}$. We first note that $\boldsymbol{z}_t(\bullet)$ in (A.1) can be expressed as a linear function of $\boldsymbol{\alpha}$. To show this, we note that the kth component of the vector $\boldsymbol{z}_t$ can be written as follows:

$$[\boldsymbol{z}_t]_k = v_t - \Big(l_t^*(\beta^k) \; v_t^*(\beta^k) \; v_t^{**}(\beta^k, \tau^k) \Big) \begin{pmatrix} \alpha_0^k \\ \alpha_1^k \\ \alpha_2^k \end{pmatrix} ,$$

with $l_t^*(\beta^k) \doteq 1 + \beta^k l_{t-1}^*(\beta^k), v_t^*(\beta^k) \doteq y_{t-1}^2 + \beta^k v_{t-1}^*(\beta^k)$ and $v_t^{**}(\beta^k, \tau^k) \doteq (\tau^k - y_{t-1})^2 \mathbb{I}\{y_{t-1} < \tau^k\} + \beta^k v_{t-1}^{**}(\beta^k, \tau^k)$ where l_0^*, v_0^*, v_0^{**} are set to zero. Let us now regroup the recursive values into a $K \times 3K$ matrix C_t as follows:

$$C_t \doteq \begin{pmatrix} l_t^*(\beta^1) & 0 & \cdots & 0 & v_t^*(\beta^1) & 0 & \cdots & 0 & v_t^{**}(\beta^1, \tau^1) & 0 & \cdots & 0 \\ 0 & l_t^*(\beta^2) & 0 & \vdots & 0 & v_t^*(\beta^2) & 0 & \vdots & 0 & v_t^{**}(\beta^2, \tau^2) & 0 & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & l_t^*(\beta^K) & 0 & \cdots & 0 & v_t^*(\beta^K) & 0 & \cdots & 0 & v_t^{**}(\beta^K, \tau^K) \end{pmatrix} .$$

It is straightforward to show that $\boldsymbol{z}_t = v_t \boldsymbol{\iota}_K - C_t \boldsymbol{\alpha}$ and since $z_t \doteq \boldsymbol{e}_t' \boldsymbol{z}_t$ we get $z_t = v_t - \boldsymbol{e}_t' C_t \boldsymbol{\alpha}$. Then, by defining the vectors $\boldsymbol{z} \doteq (z_1 \cdots z_T)'$ and $\boldsymbol{v} \doteq (v_1 \cdots v_T)'$ as well as the $T \times 3K$ matrix C whose tth row is $\boldsymbol{e}_t' C_t$, we end up with $\boldsymbol{z} = \boldsymbol{v} - C\boldsymbol{\alpha}$ which is the desired linear expression for $\boldsymbol{z}$. The proposal density to sample $\boldsymbol{\alpha}$ is obtained by combining the approximate likelihood (A.2) and the prior density by Bayes' update:

$$q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha} | \widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \boldsymbol{\nu}, \boldsymbol{s}, \boldsymbol{y}) \propto \mathcal{N}_{3K}(\boldsymbol{\alpha} | \widehat{\boldsymbol{\mu}}_{\boldsymbol{\alpha}}, \widehat{\Sigma}_{\boldsymbol{\alpha}}) \mathbb{I}\{\boldsymbol{\alpha} \geqslant \mathbf{0}\} ,$$

with

$$\begin{aligned} \widehat{\Sigma}_{\boldsymbol{\alpha}}^{-1} &\doteq C' \widetilde{\Lambda}^{-1} C + \Sigma_{\boldsymbol{\alpha}}^{-1} \\ \widehat{\boldsymbol{\mu}}_{\boldsymbol{\alpha}} &\doteq \widehat{\Sigma}_{\boldsymbol{\alpha}} \left(C' \widetilde{\Lambda}^{-1} \boldsymbol{v} + \Sigma_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\alpha}} \right) , \end{aligned}$$

where $\widetilde{\Lambda} \doteq \mathrm{diag}(\{2\boldsymbol{e}_t' \boldsymbol{h}_t^2(\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\tau})\}_{t=1}^T)$, the value $\widetilde{\boldsymbol{\alpha}}$ being the previous draw of $\boldsymbol{\alpha}$ in the Metropolis–Hasting (M–H) sampler. A candidate $\boldsymbol{\alpha}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{ \frac{p(\boldsymbol{\alpha}^\star, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, P | \boldsymbol{y})}{p(\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, P | \boldsymbol{y})} \frac{q_{\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\alpha}} | \boldsymbol{\alpha}^\star, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y})}{q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^\star | \widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y})}, 1 \right\} .$$

A.1.2. Generating $\boldsymbol{\beta}$. The function $z_t(\bullet)$ in (A.1) could be expressed as a linear function of $\boldsymbol{\alpha}$ but cannot be expressed as a linear function of $\boldsymbol{\beta}$. To overcome this problem, we linearize the vector $z_t(\boldsymbol{\beta})$ by a first order Taylor expansion at point $\widetilde{\boldsymbol{\beta}}$:

$$z_t(\boldsymbol{\beta}) \simeq z_t(\widetilde{\boldsymbol{\beta}}) + \left.\frac{\mathrm{d}z_t}{\mathrm{d}\boldsymbol{\beta}'}\right|_{\beta=\widetilde{\beta}} \times (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}),$$

where $\widetilde{\boldsymbol{\beta}}$ is the previous draw of $\boldsymbol{\beta}$ in the M–H sampler. Furthermore, let us define the following:

$$\boldsymbol{r}_t \doteq z_t(\widetilde{\boldsymbol{\beta}}) + G_t\widetilde{\boldsymbol{\beta}} \quad ; \quad G_t \doteq -\left.\frac{\mathrm{d}z_t}{\mathrm{d}\boldsymbol{\beta}'}\right|_{\beta=\widetilde{\beta}}, \tag{A.3}$$

where the $K \times K$ matrix G_t can be computed by the recursion $G_t \doteq v_{t-1}I_K - Z_{t-1} + G_{t-1}\widetilde{\boldsymbol{\beta}}$ where Z_{t-1} is a $K \times K$ diagonal matrix with $z_{t-1}(\widetilde{\boldsymbol{\beta}})$ in its diagonal, I_K is a $K \times K$ identity matrix and G_0 is a $K \times K$ matrix of zeros. From (A.3) we get $z_t \simeq \boldsymbol{r}_t - G_t\boldsymbol{\beta}$ and the approximation for z_t is obtained as $z_t \simeq r_t - \boldsymbol{e}_t' G_t \boldsymbol{\beta}$ where $r_t \doteq \boldsymbol{e}_t'\boldsymbol{r}_t$. Let us now define the vector $\boldsymbol{r} \doteq (r_1 \cdots r_T)'$ as well as the $T \times K$ matrix G whose tth row is $\boldsymbol{e}_t' G_t$. It turns out that $\boldsymbol{z} \simeq \boldsymbol{r} - G\boldsymbol{\beta}$ thus we can approximate the exponential in (A.2). The proposal density to sample $\boldsymbol{\beta}$ is obtained by combining this approximation with the prior density by Bayes' update:

$$q_\beta(\boldsymbol{\beta}|\boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y}) \propto \mathcal{N}_K(\boldsymbol{\beta}|\widehat{\boldsymbol{\mu}}_\beta, \widehat{\Sigma}_\beta)\,\mathbb{I}\{\boldsymbol{\beta} \geqslant \boldsymbol{0}\},$$

with

$$\begin{aligned} \widehat{\Sigma}_\beta^{-1} &\doteq G'\widetilde{\Lambda}^{-1}G + \Sigma_\beta^{-1} \\ \widehat{\mu}_\beta &\doteq \widehat{\Sigma}_\beta\left(G'\widetilde{\Lambda}^{-1}\boldsymbol{r} + \Sigma_\beta^{-1}\boldsymbol{\mu}_\beta\right), \end{aligned}$$

where $\widetilde{\Lambda} \doteq \mathrm{diag}(\{2\boldsymbol{e}_t'\boldsymbol{h}_t^2(\boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\tau})\}_{t=1}^T)$. A candidate $\boldsymbol{\beta}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\alpha}, \boldsymbol{\beta}^\star, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}|\boldsymbol{y})}{p(\boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}\,|\,\boldsymbol{y})}\frac{q_\beta(\widetilde{\boldsymbol{\beta}}|\boldsymbol{\alpha}, \boldsymbol{\beta}^\star, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y})}{q_\beta(\boldsymbol{\beta}^\star|\boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y})}, 1\right\}.$$

A.1.3. Generating $\boldsymbol{\tau}$. As for $\boldsymbol{\beta}$, we linearize the vector $z_t(\boldsymbol{\tau})$ by a first order Taylor expansion at point $\widetilde{\boldsymbol{\tau}}$, the previous draw of $\boldsymbol{\tau}$ in the M–H sampler. In this case, $\boldsymbol{r}_t \doteq z_t(\widetilde{\boldsymbol{\tau}}) + G_t\widetilde{\boldsymbol{\tau}}$ where the $K \times K$ matrix G_t is computed by the recursion $G_t \doteq 2I_K[\boldsymbol{\alpha}_2 \odot \mathbf{I}\{y_{t-1} < \widetilde{\boldsymbol{\tau}}\} \odot (\widetilde{\boldsymbol{\tau}} - y_{t-1})] - G_{t-1}\widetilde{\boldsymbol{\tau}}$ where I_K is a $K \times K$ identity matrix and G_0 is a $K \times K$ matrix of zeros. It turns out that $\boldsymbol{z} \simeq \boldsymbol{r} - G\boldsymbol{\tau}$, thus we can approximate the exponential in (A.2). The proposal density to sample $\boldsymbol{\tau}$ is obtained by combining this approximation with the prior density by Bayes' update:

$$q_\tau(\boldsymbol{\tau}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \widetilde{\boldsymbol{\tau}}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y}) \propto \mathcal{N}_K(\boldsymbol{\tau}|\widehat{\boldsymbol{\mu}}_\tau, \widehat{\Sigma}_\tau)\,\mathbb{I}\{\boldsymbol{\tau}_{\min} \leqslant \boldsymbol{\tau} \leqslant \boldsymbol{0}\},$$

with

$$\begin{aligned} \widehat{\Sigma}_\tau^{-1} &\doteq G'\widetilde{\Lambda}^{-1}G + \Sigma_\tau^{-1} \\ \widehat{\mu}_\tau &\doteq \widehat{\Sigma}_\tau\left(G'\widetilde{\Lambda}^{-1}\boldsymbol{r} + \Sigma_\tau^{-1}\boldsymbol{\mu}_\tau\right), \end{aligned}$$

where $\widetilde{\Lambda} \doteq \mathrm{diag}(\{2\boldsymbol{e}_t'\boldsymbol{h}_t^2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \widetilde{\boldsymbol{\tau}})\}_{t=1}^T)$. A candidate $\boldsymbol{\tau}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}^\star, \boldsymbol{\varpi}, \nu, \boldsymbol{s}|\boldsymbol{y})}{p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \widetilde{\boldsymbol{\tau}}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}|\boldsymbol{y})}\frac{q_\tau(\widetilde{\boldsymbol{\tau}}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}^\star, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y})}{q_\tau(\boldsymbol{\tau}^\star|\boldsymbol{\alpha}, \boldsymbol{\beta}, \widetilde{\boldsymbol{\tau}}, \boldsymbol{\varpi}, \nu, \boldsymbol{s}, \boldsymbol{y})}, 1\right\}.$$

A.2. Generating $\boldsymbol{\varpi}$

The components of $\boldsymbol{\varpi}$ are independent a posteriori and the full conditional posterior of ϖ_t is obtained as follows:

$$p(\varpi_t|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \nu, \boldsymbol{s}, \boldsymbol{y}) \propto \varpi_t^{-\frac{(\nu+3)}{2}} \exp\left[-\frac{b_t}{\varpi_t}\right],$$

which is the kernel of an Inverted Gamma density with parameters $(\nu + 1)/2$ and $b_t \doteq (y_t^2/(\varrho \boldsymbol{e}_t' \boldsymbol{h}_t(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau})) + \nu)/2$ where $\varrho \doteq (\nu - 2)/\nu$.

A.3. Generating ν

Draws from $p(\nu|\boldsymbol{\varpi})$ are made by optimized rejection sampling from a translated Exponential source density. The result in Deschamps (2006) can be used without modifications.

A.4. Generating $\boldsymbol{s}$ and P

The results in Chib (1996) can be used without modifications.

A.5. Computational details

The MCMC scheme is implemented in R (R Development Core Team, 2007), version 2.6.1, with some subroutines written in C. The estimation of the MSTGJR model takes about 15 minutes on a Genuine Intel ® CPU T2400 1.83 Mhz processor. Moreover, the validity of the algorithm as well as the correctness of the computer code are verified using the following methodology. We sample $\Theta \doteq (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \nu, P, \boldsymbol{\varpi}, \boldsymbol{s})$ from a proper joint prior and generate some passes of the M–H algorithm; at each pass, we simulate the dependent variable $\boldsymbol{y}$ from the full conditional $p(\boldsymbol{y}|\Theta)$ which is given by the conditional likelihood. This way, we draw a sample from the joint density $p(\boldsymbol{y}, \Theta)$. If the algorithm is correct, the resulting replications of Θ should reproduce the prior. The Kolmogorov-Smirnov empirical density test does not reject this hypothesis at the 1% level.