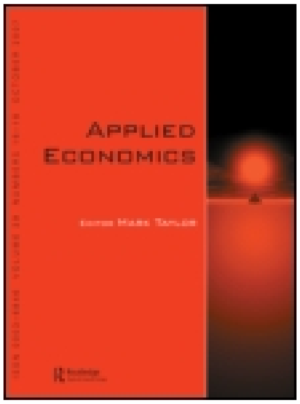

Applied Economics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/raec20

MCMC-based estimation of Markov Switching ARMA-GARCH models

Jan S. Henneke [a b], Svetlozar T. Rachev [a c], Frank J. Fabozzi [d] & Metodi Nikolov [e]

[a] University of Karlsruhe and KIT, Germany
[b] WestLB AG, London, UK
[c] University of California, Santa Barbara, CA, USA
[d] Yale School of Management, New Haven, CT, USA
[e] Finanalytica, FinAnalytica Inc., USA

Published online: 11 Mar 2009.

Applied Economics, 2011, **43**, 259–271

MCMC-based estimation of Markov Switching ARMA–GARCH models

Jan S. Henneke[a,b], Svetlozar T. Rachev[a,c], Frank J. Fabozzi[d,*] and Metodi Nikolov[e]

[a]*University of Karlsruhe and KIT, Germany*
[b]*WestLB AG, London, UK*
[c]*University of California, Santa Barbara, CA, USA*
[d]*Yale School of Management, New Haven, CT, USA*
[e]*Finanalytica, FinAnalytica Inc., USA*

Regime switching models, especially Markov Switching (MS) models, are regarded as a promising way to capture nonlinearities in time series. Combining the elements of MS models with full Autoregressive Moving Average–Generalized Autoregressive Conditional Heteroskedasticity (ARMA–GARCH) models poses severe difficulties for the computation of parameter estimators. Existing methods can become completely unfeasible due to the full path dependence of such models. In this article, we demonstrate how to overcome this problem. We formulate a full MS–ARMA–GARCH model and its Bayes estimator. This facilitates the use of Markov Chain Monte Carlo methods and allows us to develop an algorithm to compute the Bayes estimator of the regimes and parameters of our model. The approach is illustrated on simulated data and with returns from the New York Stock Exchange (NYSE). Our model is then compared to other approaches and clearly proves to be advantageous.

I. Introduction

A central property of economic time series, common to many financial time series, is that their volatility varies over time. Describing the volatility of an asset is a key issue in financial economics. Asset pricing depends on the expected volatility (covariance structure) of the returns and some derivatives depend solely on the correlation structure of their underlyings.

The most popular class of models for time-varying volatility is represented by Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-type models. In practical applications, the estimated GARCH models usually imply a very high level of persistence in volatility. This leads to the Integrated GARCH (IGARCH) model of Engle and Bollerslev (1986), where the process for volatility incorporates a unit root.

But what if the data actually stem from stationary processes that differ in their parameters? This question stirred research about structural breaks in stochastic processes. It turned out that structural breaks can account for a part of the high persistence, thus accounting for these breaks can disentangle the persistence that stems from changes in the structure and the one implied by the estimated GARCH model. Maekawa *et al.* (2005) demonstrate that most of the Tokyo stock return datasets possess volatility persistence and in many cases it is a consequence of structural breaks in the GARCH process. Rapach and Strauss (2005) report significant evidence for

*Corresponding author. E-mail: frank.fabozzi@yale.edu

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2011 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/00036840802552379

structural breaks in the unconditional variance of exchange rate returns. Smith (2002) finds strong evidence of structural breaks in GARCH models of the US index returns, foreign exchange rates and individual stock returns. He concludes that standard diagnostic tests are no substitute for structural break tests and that the results suggest that more attention needs to be given to structural breaks when building and estimating GARCH models.

Another approach to this problem would be to describe changes in the parameters endogenously with a Markov Switching (MS) model. These models were introduced to the econometric mainstream in the seminal article by Hamilton (1989). The difference is that the process can leave a state (parameter set) and returns with a positive probability. Assume that a process has a 'normal' state and several other states with higher or lower volatilities. A structural break model will base its parameter estimates only on the data between changes in the structure and discards the rest of the data. In such a scenario, an MS model will retrieve much better estimates for the 'normal' state because it operates on a much larger dataset. In this case, the MS model yields a superior fit and, more importantly, a better forecasting performance.

Market Switching models are being applied to analyse various markets. For the successful application of MS models, it is crucial to have reliable parameter estimators. In econometrics, the usual route to derive parameter estimates is to choose the Maximum Likelihood (ML) approach. However this approach becomes computationally unfeasible for MS–Autoregressive Moving Average–Generalized Autoregressive Conditional Heteroskedasticity (MS-ARMA–GARCH) models and researchers such as Cai (1994) and Hamilton and Susmel (1994) have dismissed these models as too untractable. Instead, they use low order MS–Autoregressive–Autoregressive Conditional Heteroskedasticity (AR-ARCH) models for which they derived estimators.

In this article, we develop an algorithm for the estimation of the parameters of a full MS–ARMA–GARCH model. For this we chose the Bayesian framework, because this enables the application of Markov Chain Monte Carlo (MCMC) methods which are powerful tools for the numerical computation of integrals. We proceed as follows. In Section II, we present the model specification. In Section III, we present our estimation algorithm for the proposed MS–ARMA–GARCH model. Thereafter, we evaluate our algorithm in Section IV on empirical data and compare it against competing model specifications. Section V concludes our article.

II. MS–ARMA–GARCH Model

We consider the straightforward extension of Hamilton's original MS model to a MS–ARMA–GARCH model. The latter model is simply an ARMA–GARCH model whose parameters are controlled by latent regimes or states. These states are themselves random and are assumed to follow a discrete $\mathcal{S}$-dimensional ergodic Markov chain $\{S_t\}_{t\in\mathbb{N}}$ defined on the discrete state space $\{1, 2, \ldots, \mathcal{S}\}$ whose transition probabilities $\pi_{i,j} - P(S_t = j|S_{t-1} = i)$ are collected in a transition probability matrix Π. The MS–ARMA–GARCH model whose parameters are dependent on the state of this Markov chain is defined in Equations 1a–1c.

$$y_t = c_{S_t} + \sum_{i=1}^{r} \phi_i(S_t) \cdot y_{t-i} + \varepsilon_t + \sum_{j=1}^{m} \psi_j(S_t) \cdot \varepsilon_{t-j} \quad (1a)$$

$$h_t = \omega_{S_t} + \sum_{i=1}^{q} \alpha_i(S_t) \cdot \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j(S_t) \cdot h_{t-j} \quad (1b)$$

$$\varepsilon_t = \sqrt{h_t} \cdot u_t \quad u_t \sim N(0, 1) \quad (1c)$$

Francq and Zakoian (2001) consider an MS–ARMA model corresponding to Equation 1a and Francq and Zakoian (2002) discuss the stationarity properties[1] and moment estimators for an MS–GARCH as in Equation 1b. In the next section, we develop an MCMC algorithm to compute the Bayes estimator for the parameters of our model.

III. Estimating the Model Parameters

In order to compute the Bayes estimator for the parameters of the proposed model, we need to specify the full Bayesian statistical model. We will choose conjugate priors wherever possible and normal priors with adequate hyperparameters in the other cases. Since we will work on large datasets, we do not consider the choice of the prior distribution as a critical issue and rely on the asymptotic efficiency of the Bayes estimator. The complete parameter space for our model is given by

$$\Theta = \{\Pi \times \Theta_{\mathrm{ARMA}} \times \Theta_{\mathrm{GARCH}}\}$$

Note that this does include all nonstationary specifications of the parameters. We will impose stationarity

[1] See Appendix B for details.

through the Metropolis–Hastings (MH)-step in which we will restrict the acceptance of a parameter to a subset on Θ.

$$\begin{aligned}\Theta_{\text{ARMA}_1} &\sim N\left(\mu_{\text{ARMA}_1}, \Sigma_{\text{ARMA}_1}\right) \cdot \mathbf{1}_S\left(\theta_{\text{ARMA}_1}\right)\\ \Theta_{\text{GARCH}_1} &\sim N\left(\mu_{\text{GARCH}_1}, \Sigma_{\text{GARCH}_1}\right) \cdot \mathbf{1}_S(\left(\theta_{\text{GARCH}_1}\right)\\ &\vdots\\ \Theta_{\text{ARMA}_S} &\sim N\left(\mu_{\text{ARMA}_S}, \Sigma_{\text{ARMA}_S}\right) \cdot \mathbf{1}_S\left(\theta_{\text{ARMA}_S}\right)\\ \Theta_{\text{GARCH}_S} &\sim N\left(\mu_{\text{GARCH}_S}, \Sigma_{\text{GARCH}_S}\right) \cdot \mathbf{1}_S\left(\theta_{\text{GARCH}_S}\right)\\ \Pi &\sim \text{Dirichlet } (\alpha_1, \ldots, \alpha_S)\end{aligned}$$

where the indicator functions $\mathbf{1}_S(\theta) = 1$ for a parameter set which is stationary, 0 otherwise. The complete prior distribution for θ is

$$[\theta] = \text{Dirichlet } (\alpha_1, \ldots, \alpha_S) \times \sum_{i=1}^{S} N\left(\mu_{\text{ARMA}_i}, \Sigma_{\text{ARMA}_i}\right) \times N\left(\mu_{\text{GARCH}_i}, \Sigma_{\text{GARCH}_i}\right) \quad (2)$$

To compute the likelihood of the model, the parameter vector θ is augmented[2] with the states $S_{[1,T]}$. We can then compute the posterior of all unobservable quantities and not only get estimates for the model parameters $\hat{\theta}$, but also estimates for the states $\hat{S}_{[1,T]}$.

$$\begin{aligned}p\left(\theta, S_{[1,T]}|y\right) &\propto f\left(y|\theta, S_{[1,T]}\right)p\left(\theta, S_{[1,T]}\right)\\ &\propto f\left(y|\theta, S_{[1,T]}\right)p\left(S_{[1,T]}|\theta\right)p(\theta)\end{aligned}$$

We can see that computing the posterior mean is a difficult task. In fact, our major problem to be solved is to compute or approximate the posterior mean for the full statistical model.

$$\left\{\hat{\theta}, \hat{S}_{[1,T]}\right\} = E\left[\theta, S_{[1,T]}|Y = y\right] \quad (3)$$

We will now construct an MCMC algorithm that produces a series of samples:

$$\left\{\theta_1^{(g)}, \ldots, \theta_m^{(g)}, S_{[1,T]}^{(g)}\right\} \quad g \in \mathbb{N}, \quad m = \dim(\Theta)$$

which will converge to the joint posterior distribution. To construct the MCMC algorithm, we will use a hybrid method consisting of both Gibbs steps and MH steps.[3]

Implementing the MCMC algorithm

To sample from the individual full conditional posterior distributions, we need to choose adequate prior distributions for the parameters. We use the priors as proposed above. If we can obtain an analytic expression for the full conditional posterior density, then we use a Gibbs step to obtain the sample since an MH step is computationally more intensive. Otherwise, we can just use a rather diffuse normal prior[4] because its influence will vanish on samples of the size that we consider. Therefore, we use normal priors for all ARMA and GARCH coefficients. The steps in the MCMC algorithm are described as follows.

Sampling the transition probabilities. The posterior distribution of $\pi_{i,j}$ is given by

$$p\left(\pi_{i,j}|y, S_{[1,T]}, \Theta\backslash\pi_{i,j}\right) \propto p\left(\pi_{i,j}\right)p\left(S_{[1,T]}, y|\Theta\right)$$

Since S_t is independent of y, this is

$$p\left(\pi_{i,j}|y, S_{[1,T]}, \Theta\backslash\pi_{i,j}\right) \propto p\left(\pi_{i,j}\right)p\left(S_{[1,T]}|\Theta\right)$$

Let $\eta_{i,j}$ be the cumulated number of transitions from state i to state j in the current sample $S_{[1,T]}^{(g)}$. Then we can write this as

$$\begin{aligned}p\left(S_{[1,T]}|\Theta\right) &= \prod_{t=1}^{T} p(S_{t+1}|S_t, \Theta)\\ &= \left(\pi_{1,1}\right)^{\eta_{1,1}}\left(\pi_{1,2}\right)^{\eta_{1,2}}\cdots\left(\pi_{1,\mathcal{S}}\right)^{\eta_{1,\mathcal{S}}}\left(\pi_{2,1}\right)^{\eta_{2,1}}\cdots\end{aligned} \quad (4)$$

for each row of Π, $\pi_s = (\pi_{s,1}, \ldots, \pi_{s,\mathcal{S}})$. This is proportional to the density from a Dirichlet distribution. A conjugate prior would thus be a Dirichlet distribution with the hyperparameters $\alpha_s = (\alpha_{s,1}, \ldots, \alpha_{s,\mathcal{S}})'$:

$$f(x|\alpha_s) = \frac{1}{B(\alpha_s)}\prod_{i=1}^{\mathcal{S}} x_i^{\alpha_{s,i}-1} \qquad B(\alpha_s) = \frac{\Pi_{i=1}^{\mathcal{S}}\Gamma(\alpha_{s,i})}{\Gamma\left(\Sigma_{i=1}^{\mathcal{S}}\alpha_{s,i}\right)}$$

The posterior is then again a Dirichlet distribution with the parameters $\alpha + \eta$. We therefore obtain a sample of the transition probabilities from state s to all others by generating a draw from

$$P(\Pi_s|\alpha_s) = \text{Dirichlet } (\alpha_{s,1} + \eta_{s,1}, \ldots, \alpha_{s,\mathcal{S}} + \eta_{s,\mathcal{S}})$$

In the next step, we need to obtain a sample of the states. We will follow the single move scheme suggested by Carlin *et al.* (1992).

Sampling $S_{[1,T]}$. In this step of the MCMC algorithm, we obtain a sample from the distribution of the

[2] See Tanner and Wong (1987).

[3] For a more elaborate discussion of the Gibbs and MH algorithm in the context of our problem, see Henneke *et al.* (2006). For the general regularity conditions on the MH and Gibbs algorithm, see Robert and Casella (1999) and Tierny (1994); for ARMA models in particular, see Chib and Greenberg (1994).

[4] Choose a mean and σ of 0.5 which centers the ARMA–GARCH parameters on [0,1] and leaves all stationary values in an area of roughly one SD. This makes the influence of the prior virtually zero.

entire Markov chain $S_{[1,T]}$. One possibility is to compute the measure $p(S_t|y, \Theta)$, but because of the path dependence of the likelihood $p(y|S_{[1,T]}, \Theta)$ the time complexity is in $\mathcal{O}(\mathcal{S}^T)$ and therefore computationally not feasible. The single move procedure breaks down this step into a Gibbs cycle of T consecutive draws from the conditional distribution of the state at a single point in time, conditional on all other states. This is done as follows.

First, we compute the measure $p(S_t|\{S\backslash S_t\}, \Theta, y)$. Then we write $\{S\backslash S_t\}$ as $S_{\neq t}$, $S_{[1,T]}$ as S and omit to explicitly condition on Θ.

$$p(S_t|S_{\neq t}, y) = \frac{p(S_t, S_{\neq t}, y)}{p(y, S_{\neq t})} = \frac{p(y|S)\cdot p(S)}{p(y|S_{\neq t})\cdot p(S_{\neq t})} = \frac{p(y|S)\cdot p(S|S_{\neq t})}{p(y|S_{\neq t})}$$

$p(y|S)$ is computed easily. With a given sample of S, this is simply the likelihood of the model. $p(S_t|S_{\neq t})$ is only dependent on S_{t-1} and S_{t+1} due to the Markov property of the chain.

$$p(S_t = i|S_{\neq t}) = p(S_t = i|S_{t-1}, S_{t+1}) = \frac{\pi_{l,i}\cdot\pi_{i,k}}{\Sigma_{i=1}^{\mathcal{S}}\pi_{l,i}\cdot\pi_{i,k}}$$

with $S_{t-1} = l$, $S_{t+1} = k$ and $\pi_{i,j}$ the respective transition probabilities from Π. Since for all $S_t = i$, $i \in \{1, \ldots, \mathcal{S}\}$, $p(y|S_{\neq t})$ is constant, we write

$$p(S_t = i|y, S_{\neq t}) \propto p(y|S_{t=i}, S_{\neq t})\cdot p(S_t = i|S_{\neq t})$$

Because $p(S_t|y, S_{\neq t})$ is a probability measure, we can now compute it as

$$p(S_t = i|y, S_{\neq t}) = \frac{p(y|S_t = i, S_{\neq t})\cdot p(S_t = i|S_{\neq t})}{\Sigma_{i=1}^{\mathcal{S}} p(y|S_t = i, S_{\neq t})\cdot p(S_t = i|S_{\neq t})}$$

One sample of S is thus obtained by cycling through the following steps for each $t \in \{1, \ldots, T\}$, starting with $t = 1$:

- Compute the distribution $p(S_t = i|S_{\neq t}, y)$ on $\{1, \ldots, \mathcal{S}\}$.
- Draw S_t from this distribution.
- Update S with this value.

Sampling the ARMA–GARCH parameters

We will use an MH step to obtain samples from the full conditional posterior distributions of these parameters. For the speed of the convergence it is crucial to select an adequate proposal distribution.[5] We will therefore exploit all the knowledge we have about the full conditional posterior.

Sampling the parameters of the conditional mean. The likelihood conditional on the augmented parameter space and observed data is

$$\begin{aligned} &p\big(c_i|\Theta\backslash c_i, S_{[1,T]}, y\big) \\ &\quad\propto f\big(y|\Theta, S_{[1,T]}\big)p(c) \\ &\quad\propto \prod_{t=1}^{T}\frac{1}{\sqrt{2\pi h_t}}\exp\left\{-\frac{\big(y_t - c_{S_t} - \phi_{S_t}y_{t-1} - \psi_{S_t}\varepsilon_{t-1}\big)^2}{2h_t}\right\} \\ &\quad\times p\big(c_{S_t}\big)\end{aligned}$$

Treating all other parameters as constants, this is only a function of c_i, and with $\mathcal{C}_{t,S_t} = y_t - \phi_{S_t}y_{t-1} - \psi_{S_t}\varepsilon_{t-1}$ we rewrite it as

$$\begin{aligned} &\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi h_t}}\exp\left\{-\frac{\big(\mathcal{C}_{t,S_t} - c_{S_t}\big)^2}{2h_t}\right\}p\big(c_{S_t}\big) \\ &\quad= \prod_{t=1}^{T}\frac{1}{\sqrt{2\pi h_t}}\exp\left\{-\frac{\mathcal{C}_{t,S_t}^2 - 2\mathcal{C}_{t,S_t}\cdot c_{S_t} + c_{S_t}^2}{2h_t}\right\}p\big(c_{S_t}\big) \\ &\quad\propto \exp\left\{-\left(c_i\cdot\sum_{t=1}^{T}\frac{-\mathcal{C}_{t,S_t}\mathbf{1}_{[S_t=i]}}{h_t} + c_i^2\sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}}{2h_t}\right)\right\}p(c_i)\end{aligned}$$

This has the form of a normal distribution with

$$\mu_{c_i} = \sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}\mathcal{C}_{t,S_t}}{h_t}\cdot\left(\sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}}{h_t}\right)^{-1}, \quad \sigma_{c_i}^{-2} = \sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}}{h_t}$$

Hence we choose $N(\mu_c, \sigma_c^2)$ as the proposal distribution. For the other parameters we can proceed in a similar fashion and get

$$\begin{aligned} \mu_{\phi_i} &= \sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}\mathcal{C}_t y_{t-1}}{h_t}\cdot\left(\sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}y_{t-1}^2}{h_t}\right)^{-1}, \\ \sigma_{\phi_i}^{-2} &= \sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}y_{t-1}^2}{h_t} \\ \mu_{\psi_i} &= \sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}\mathcal{C}_t\varepsilon_{t-1}}{h_t}\cdot\left(\sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}\varepsilon_{t-1}^2}{h_t}\right)^{-1}, \\ \sigma_{\psi_i}^{-2} &= \sum_{t=1}^{T}\frac{\mathbf{1}_{[S_t=i]}\varepsilon_{t-1}^2}{h_t}\end{aligned}$$

Sampling the GARCH coefficients. As shown by Engle and Bollerslev (1986), a GARCH(p, q) process is expressed as an ARMA(l, s) process of

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{l}\big(\alpha_j + \beta_j\big)\varepsilon_{t-j}^2 + \tilde{w}_t - \sum_{j=1}^{s}\beta_j\tilde{w}_{t-j}$$

[5] See Robert (1994), Robert and Casella (1999) or Henneke *et al.* (2006).

with $\alpha_j = 0$ for $j > p$, $\beta_j = 0$ for $j > q$, $l = \min(p, q)$, $s = q$, and

$$\tilde{w}_t := \varepsilon_t^2 - \sigma_t^2 = \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right)\sigma_t^2 = (\chi^2(1) - 1)\sigma_t^2$$

The conditional mean of $\tilde{w}_t$ is $E[\tilde{w}_t|\mathcal{F}_{t-1}] = 0$, and the conditional variance is $\mathrm{Var}(\tilde{w}_t|\mathcal{F}_{t-1}) = 2\sigma_t^4$. Nakatsuma (1998) suggests replacing this $\tilde{w}_t$ with $w^* \sim N(0, 2\sigma_t^4)$. Then we have an auxiliary ARMA model for the squared errors ε_t^2

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{l} (\alpha_j + \beta_j)\varepsilon_{t-j}^2 + w_t - \sum_{j=1}^{s} \beta_j w_{t-j},$$
$$w_t \sim N(0, 2\sigma_t^4) \tag{5}$$

Rewriting this expression and factoring out β_j, we get

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{l} \alpha_j \varepsilon_{t-j}^2 + w_t + \sum_{j=1}^{s} \beta_j \left(\varepsilon_{t-j}^2 - w_{t-j}\right)$$

$$w_t = \varepsilon_t^2 - \omega - \sum_{j=1}^{l} \alpha_j \varepsilon_{t-j}^2 - \sum_{j=1}^{s} \beta_j \underbrace{\left(\varepsilon_{t-j}^2 - w_{t-j}\right)}_{=:v_t} \sim N(0, 2\sigma_t^4)$$

In our GARCH(1,1), setting $l = 1$ and $s = 1$ we arrive at a posterior distribution for ω as follows:

$$p(\omega_1|\Theta\backslash\omega_1, S, y)$$
$$\propto \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi 2h_t^2}} \exp\left\{-\frac{\left(\varepsilon_t^2 - \omega_{S_t} - \alpha_{S_t}\varepsilon_{t-1}^2 - \beta_{S_t} v_{t-1}\right)^2}{4h_t^2}\right\}$$
$$\times p(\omega_1)$$

As said before, we write $\mathcal{C}_t = \varepsilon_t^2 - \mathbf{1}_{[S_t=1]}\omega_2 - \alpha_{S_t}\varepsilon_{t-1}^2 - \beta_{S_t} v_{t-1}$ and obtain

$$p(\omega_1|\Theta\backslash\omega_1, S, y)$$
$$\propto \exp\left\{-\left(\omega_1 \sum_{t=1}^{T} \frac{-\mathbf{1}_{[S_t=1]}\mathcal{C}_t}{2h_t^2} + \omega^2 \sum_{t=1}^{T} \frac{\mathbf{1}_{[S_t=1]}}{4h_t^2}\right)\right\} p(\omega_1)$$

Our proposal distribution for ω_1 is therefore $N(\mu_{\omega_1}, \sigma_{\omega_1}^2)$ with

$$\mu_{\omega_1} = \sum_{t=1}^{T} \frac{\mathbf{1}_{[S_t=1]}\mathcal{C}_t}{2h_t^2} \cdot \left(\sum_{t=1}^{T} \frac{\mathbf{1}_{[S_t=1]}}{2h_t^2}\right)^{-1},$$
$$\sigma_{\omega_1}^{-2} = \sum_{t=1}^{T} \frac{\mathbf{1}_{[S_t=1]}}{2h_t^2}$$

Proposal distributions for α and β are obtained analogue to these results.[6]

Estimating the parameters of the t-distributed innovations

In the model where the innovations are Student t-distributed, the algorithm changes slightly. The MH steps essentially stay the same, but we will adjust the proposal distribution. We also have to estimate the degree of freedom parameter v_{S_t}.

Sampling $\boldsymbol{v}$. We follow Jacquier *et al.* (2004) and choose a discrete flat prior for v, whereas Geweke (1993) uses a continuous prior to estimate the degree of freedom in Student t-linear models. The posterior distribution of v is proportional to the product of t-distribution ordinates:

$$p\left(v|\Theta\backslash v, S_{[1,T]}, y\right) \propto p(v)p\left(y|\Theta, S_{[1,T]}\right)$$
$$\propto \prod_{t=1}^{T} \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})}\left(1 + \frac{\hat{e}_t^2}{h_t v}\right)^{-\left(\frac{v+1}{2}\right)}$$

where

$$\hat{e}_t = y_t - \sum_{i=1}^{r} \phi_i(S_t) y_{t-i} - \sum_{i=1}^{m} \psi_i(S_t)\varepsilon_{t-i}$$

Theoretically v is in $\mathbb{N}^+$, but we choose a flat prior on $\{3, \ldots, 40\}$. The posterior distribution can then be calculated analytically as follows:

$$\tilde{p}\left(v|\Theta\backslash v, S_{[1,T]}, y\right) = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{\hat{e}_t^2}{h_t v}\right)^{-\left(\frac{v+1}{2}\right)} \tag{6}$$

Then for $p(v|\Theta\backslash v, S_{[1,T]}, y)$ with a flat prior on $\{3, \ldots, 40\}$, $p(v) = \frac{1}{37}$, we get:

$$p\left(v|\Theta\backslash v, S_{[1,T]}, y\right) = \frac{\tilde{p}\left(v|\Theta\backslash v, S_{[1,T]}, y\right)}{\Sigma_{i=3}^{40}\tilde{p}\left(i|\Theta\backslash i, S_{[1,T]}, y\right)} \cdot \frac{1}{37}$$

Truncating the interval of v on $\{3, \ldots, 40\}$ will not result in inaccuracies as long as the sampled values $v^{(g)}$ do not touch the boundaries. We will only choose to model data with a t-distribution if the degree of freedom is significantly smaller than 30. Above 30 it is common in the statistical literature to approximate the t-distribution with the normal since they are very close.

Proposal distributions. We make use of the fact that for the Student t-distribution, there exists a hidden mixture representation through the normal distribution. $p(x|\theta)$ is the mixture of a normal distribution

[6] h_t, which is itself dependent on the parameters, can then be computed in the MH step with the current parameter set $\theta^{(g)}$ (see also Section 'The complete algorithm').

and an inverse γ distribution (see, e.g. Robert, 1994).

$$x|z \sim N(\theta, z\sigma^2), z^{-1} \sim \gamma\left(\frac{v}{2}, \frac{v}{2}\right)$$

We can now rewrite the innovations (1c) in the model as

$$\eta_t = \sqrt{h_t} \cdot \tilde{\eta}_t \tag{7a}$$

$$\tilde{\eta}_t = \sqrt{\lambda_t} \cdot u_t \sim t(v), \quad u_t \sim N(0,1), \quad \lambda_t \sim \mathcal{IG}\left(\frac{v}{2}, \frac{v}{2}\right) \tag{7b}$$

Conditional on a sample $\lambda_{[1,T]}^{(g)}$, the normalized residuals are again normal and we can use our previously developed methods for the ARMA–GARCH parameters. The sample $\lambda_{[1,T]}^{(g)}$ is obtained in an individual Gibbs step.

$$\varepsilon_t = \left(y_t - c_{S_t} - \sum_{i=1}^{r} \phi_i(S_t) \cdot y_{t-i} - \sum_{j=1}^{m} \psi_j(S_t) \cdot \eta_{t-j}\right) \frac{1}{\sqrt{h_t \lambda_t}} \tag{8}$$

Sampling λ_t. The posterior distribution of λ_t is given by

$$p(\lambda_t|\eta_t, h_t, v) \propto \lambda_t^{-\frac{(v+3)}{2}} \exp\left\{-\left(\frac{\eta_t^2}{h_t} + v\right)\frac{1}{2\lambda_t}\right\}$$

This is proportional to the density of a χ^2 random variable and we have

$$\left(\frac{\eta_t^2}{h_t} + v\right)\frac{1}{\lambda_t} \sim \chi^2(v+1)$$

To obtain the sample λ_t, we draw from $x \sim \chi^2(v+1)$ and compute $\lambda_t = (\frac{\eta_t^2}{h_t} + v)\frac{1}{x}$. We have now developed all the necessary steps of the algorithm which we will summarize in the next section.

The complete algorithm

Our algorithm creates samples $\{\theta^{(g)}, S^{(g)}\}$ from a Markov chain which converge to the joint posterior distribution $p(\Theta, S_{[1,T]}|y)$. Now we present a precise description of the algorithm:

The sample $\theta^{(g)}$ denotes the parameter set obtained in the g-th step. The sample value $\theta^{(g+1)}$ is obtained by iterating through the following steps:

(1) Sample $\Pi^{(g+1)}$: Draw $\pi_{i,j}$ from

$$P(\Pi_s|\alpha_s) = \text{Dirichlet}\ (\alpha_{s,1} + \eta_{s,1}, \ldots, \alpha_{s,\mathcal{S}} + \eta_{s,\mathcal{S}})$$

(2) Sample $S^{(g+1)}$ by the single move procedure from

$$p\left(S_t = i|y, S_{\neq t}\right) = \frac{p\left(y|S_{[1,T]}\right) \cdot p\left(S_t = i|S_{\neq t}\right)}{\Sigma_{i=1}^{S} p\left(y|S_{[1,T]}\right) \cdot p\left(S_t = i|S_{\neq t}\right)}$$

(3) Sample $\lambda_{[1,T]}^{(g+1)}$: Compute

$$\hat{\eta}_t = y_t - c_{S_t} - \sum_{i=1}^{r} \phi_i(S_t) \cdot y_{t-i} - \sum_{j=1}^{m} \psi_j(S_t) \cdot \hat{\eta}_{t-j}$$

$$\hat{h}_t = \omega_{S_t} + \sum_{i=1}^{p} \alpha_i(S_t) \cdot \hat{\eta}_{t-i}^2 + \sum_{j=1}^{q} \beta_j(S_t) \cdot \hat{h}_{t-j}$$

with the parameters taken from the current sample $\theta^{(g)}$ and S_t from $S^{(g+1)}$. Then draw a sample $x_{[1,\ T]}$ from $\chi^2(v+1)$ with v from $\theta^{(g)}$ and compute $\lambda_t = (\frac{\eta_t^2}{h_t} + v)\frac{1}{x_t}$.

(4) Sample $v^{(g+1)}$: Use the same $\hat{\eta}_t$ and $\hat{h}_t$ from the previous step to sample from

$$p\left(v|\theta^*\backslash v, S_{[1,T]}, y\right) = \frac{\tilde{p}\left(v|\theta^*\backslash v, S^{(g+1)}, y\right)}{\Sigma_{i=3}^{40}\tilde{p}(i|\theta^*\backslash i, S^{(g+1)}, y)} \cdot \frac{1}{37}$$

with $\tilde{p}(i|\theta^*\backslash i, S^{(g+1)}, y)$ from Equation 6. θ^* is the current parameter set at $\theta^{(g)}$ updated with the parameters sampled in Steps 1–3. Using the same residuals and conditional variances from Step 3 helps to avoid problems documented by Eraker *et al.* (1998), in which v can get absorbed into the lower bound.

(5) Cycle through the ARMA–GARCH parameters: For each parameter draw from the respective proposal distribution by the scheme outlined above.

The candidate is then tried in an MH step and either accepted or rejected.

Let $\hat{\vartheta}$ denote the proposed parameter value. We update θ^* with this value and refer to this parameter set as $\hat{\theta}$. We accept $\hat{\theta}$ as the new θ^* with probability

$$\alpha_{MH}\left(\theta^*, \hat{\theta}\right) = \min\left\{\frac{p(\hat{\theta}|y, S^{(g+1)})}{p(\theta^*|y, S^{(g+1)})} \Big/ \frac{g\left(\hat{\theta}\right)}{g(\theta^*)}, 1\right\}$$

where $p(\hat{\theta}|y, S^{(g+1)})$ is the likelihood of the data y and the current sample $S^{(g+1)}$, calculated with the parameter set $\hat{\theta}$. $g(.)$ is the respective proposal density. This is done for all parameters of the ARMA–GARCH specification. The parameters which violate stationarity conditions will be rejected. This does effectively truncate the distribution on the stationary interval. Finally, we have obtained the new sample $\{\theta^{(g+1)}, S^{(g+1)}\}$.

Estimating the Hamilton–Susmel model with an MCMC algorithm

For the purpose of comparison, we will also estimate a variant of the Hamilton–Susmel model. With the method developed in this section, it is not a big step to compute the estimators for that model. In the Hamilton–Susmel model, the latent state governing the evolution of the model parameters is assumed to follow the same $\mathcal{S}$-dimensional time discrete Markov chain as in the previous model. The conditional mean of the time series $\{y_t\}$ is as in Equation 1a, the innovations and the variance process are specified as

$$\begin{aligned} h_t &= w + \sum_{i=1}^{q} \alpha_i \cdot \varepsilon_{t-i} + \sum_{j=1}^{p} \beta_j \cdot h_{t-j} \\ \varepsilon_t &= \sqrt{g_{S_t}} \cdot \sqrt{h_t} \cdot u_t \\ u_t &\sim N(0,1) \end{aligned}$$

The mean is the same as in our model and the conditional variance is an *amplified* GARCH process. The main difference between the Hamilton–Susmel model and our model is the specification of the variance. In Hamilton and Susmel (1994) the authors examine model specifications of the type AR(r)–ARCH(q), whereas we focus on GARCH specifications of the variance and ARMA specifications of the mean. All parameters of the model can be estimated through a straightforward application of the methods shown in the previous sections with exception of the amplifying parameter. Therefore, we will now discuss how the amplifying parameter g_{S_t} can be estimated within our MCMC setting.

The likelihood $f(\theta|y, S_{[1,T]})$ is given by

$$\begin{aligned} f(\theta|y, S_{[1,T]}) &= \prod_{t_1 \in \mathcal{I}_1} \frac{1}{\sqrt{2\pi g_1 h_t}} \exp\left\{-\frac{\varepsilon_{t_1}^2}{2 g_1 h_t t_1}\right\} \\ &\quad \times \cdots \times \prod_{t_S \in \mathcal{I}_S} \frac{1}{\sqrt{2\pi g_S h_t}} \exp\left\{-\frac{\varepsilon_{t_s}^2}{2 g_S h_{t_s}}\right\} \end{aligned}$$

where $\mathcal{I}_s$ is the index set containing all points in time when $S_t = s$. It is easy to see that for a certain g_s, the likelihood of g_s is proportional to

$$\begin{aligned} f(g_s|\{\theta\backslash g_s\}, y, S_{[1,T]}) &\propto \prod_{t_s \in \mathcal{I}_s} \frac{1}{\sqrt{2\pi h_t}} g^{-\frac{1}{2}} \exp\left\{-\frac{\varepsilon_t^2/(2h_t)}{g_s}\right\} \\ &\propto g_s^{-N_s/2} \exp\left\{-\frac{1}{g_s} \sum_{t \in \mathcal{I}_s} \frac{\varepsilon_t^2}{2h_t}\right\} \end{aligned}$$

where $N_s = \mathrm{card}(\mathcal{I}_s)$. This has the form of an *Inverse* γ density, which is

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left\{\frac{-\beta}{x}\right\}$$

with shape parameter α and scale parameter β. Therefore, a conjugate prior $p(g_s)$ is an inverse γ distribution with the hyperparameters α_0, β_0. The posterior distribution for g_s would then be

$$\begin{aligned} & f(g_s|\{\theta\backslash g_s\}, y, S_{[1,T]}) \\ & \quad \propto g^{-N_s/2} \exp\left\{-\frac{1}{g_s} \sum_{t \in \mathcal{I}_s} \frac{\varepsilon_t^2}{2h_t}\right\} \cdot p(g_s|\alpha_0, \beta_0) \end{aligned}$$

and hence

$$g_s \sim \mathcal{IG}\left(\frac{1}{2} N_s + \alpha_0, \sum_{t \in \mathcal{I}_s} \frac{\varepsilon_t^2}{2h_t} + \beta_0\right)$$

We specified the innovations u_t as coming from a normal distribution. We could also let them be Student t-distributed, in the estimation procedure we would then simply have to apply the result from the Section 'Proposal distributions'.

IV. Estimation Results

Model diagnostics

We assess the in-sample goodness of fit by using transformed one-step-ahead forecast errors. The standardized residuals would not be identically distributed, which prevents us from using the standard methods. We follow Kaufmann and Fruewirth-Schnatter (2000) and Haas *et al.* (2004) and compute

$$u_t = F(\hat{\varepsilon}_t|\Theta, S_{[1,T]}), \quad t = 1, \ldots, T.$$

and use various standard diagnostics on the transformed residuals $\hat{u}_t = \Phi^{-1}(u_t)$, where Φ is the standard normal cumulative distribution function. In particular we examine the sample Autocorrelation Function (sample ACF), normal probability plot (QQplot) and the Lagrange Multiplier (LM) test of Engle (1982). In addition, we report the statistics for the skewness, excess kurtosis and the Jarque–Bera test statistic for $\hat{u}_t$.

Empirical results

For the purpose of comparison, we examine exactly the same dataset as Hamilton and Susmel (1994), who use weekly returns from the New York Stock

Exchange (NYSE).[7] We now provide the estimation results of our model specification and compare them to the Hamilton–Susmel model.

Hamilton and Susmel (1994) estimate the parameters for their model and compare several different parameterizations to each other. They conclude that the *t*-Switching ARCH (SWARCH)-L(3,2) specification is the most appropriate model for that data. In the notation of the Hamilton–Susmel model, it is an AR(1)–ARCH(2) process with a leverage effect and three regimes controlling the amplifier g_{S_t}. They assessed the suitability of their model mainly through the forecasting properties. They observed that the single regime GARCH models implied a very high persistence, but did a worse job forecasting the volatility than a simple constant variance model measured by mean squared errors. This is counter-intuitive, since if the variance was persistent, that would imply better forecasts with a model that accounts for this persistence.

We prefer to assess the fit through the sample ACF, the LM test and the Jarque–Bera statistic of the estimated transformed residuals. We found that a simple AR(1) specification of the conditional mean seemed to be an unlikely candidate to capture this rather complex structure. We estimated the standard ARMA(1,1)–GARCH(1,1) model with Gaussian innovations and with *t*-distributed innovations on the data.[8] The sample ACF of the residuals showed no significant autocorrelations indicating that the single regime ARMA might be enough to describe the mean of the weekly return series. In fact, as we fitted a two regime model of the mean, the results of the algorithm were very unstable in the estimated states, also an indication of no significant switching in the mean.

We then fitted an MS(1,2)–ARMA(1,1)–GARCH(1,1) specification of both models to the data and could observe that the persistence due to the GARCH component decreased. The fit of both of these models as measured by the sample ACF was still unsatisfactory.

We chose to model the conditional variance as a three-regime process and estimated such a specification of the Hamilton–Susmel model and our model on weekly NYSE returns. The resulting estimators and posterior density statistics are reported in Tables 1 and 2, respectively. We produced an MCMC chain of 50 000 samples for each parameter and discarded the first 10 000 to account for any burn-in effects. The qualitative CUSUM criteria which allows for a visual inspection of the convergence of the Markov Chain indicates a satisfactory convergence, whereas the convergence for some parameters is rejected in a nonparametric Kolmogorov–Smirnov test (Lehmann, 1975; Robert and Casella, 1999). Overall, we find that the algorithm yields a very stable and reliable result.

In both models, the third regime can be termed as the high-volatility regime and we can regard the second regime as the normal state. The first regime appears only once. Therefore, we cannot be sure whether this is a structural break in the process or a state that can govern the parameters again.

In our estimates, we can observe that β_3 is the highest value among the β's in our model, at the same time it is smaller than the β from the Hamilton–Susmel model. This shows that the high level of persistence in the variance process is attributed more to MS than to a GARCH effect.

Figures 1 and 2 compare the estimated posterior probabilities for both models. Our model is able to distinguish the different states much sharper, which naturally is a very desirable feature.

Next, we compare the sample ACF of the two models. Since the ARMA parameters are estimated to be about the same in both models, we only show the sample ACF for the squared residuals. Figures 3 and 4 show that our model captures the autocorrelation structure of the data much better. Tables 1 and 2 show that the posterior SDs of the ARMA parameters of our proposed model are less than those of the Hamilton–Susmel model. This also indicates an improvement in the specification of the variance process. The SDs for the GARCH parameters are higher for the proposed model. This is a natural consequence of our specification of the model compared to the Hamilton–Susmel model. The GARCH parameters for the Hamilton–Susmel model are estimated on the whole dataset. In our model they are only estimated for their regime. However, they compare well to posterior SDs retrieved on smaller samples,[9] even though the

[7] The stock price series used is the value-weighted portfolio of stocks traded on the NYSE contained in the Center for Research in Security Prices (CRSP) data tapes. Hamilton and Susmel use the weekly returns from Wednesday of 1 week to Tuesday of the following week. The first return is from the week ended 31 July 1962 to the week ended 29 December 1987.
[8] The observed values of the sample ACF of the residuals were extremely small. This is caused by the extreme innovation of the crash in 1987. In this model such an event would have to be regarded as an outlier. If we would not regard this value as an outlier, this indicates that the data actually stems from a distribution that has much heavier tails. The autocorrelation as a measure for dependence is then not very meaningful. Nevertheless, we computed the sample ACF on a truncated dataset, excluding this extreme event. This gives us a proxy by which we can diagnose the model fit to the rest of the data.
[9] See also Chib and Greenberg (1994) or Bauwens and Lubrano (1998).

Table 1. Estimated parameter values and posterior statistics for the modified Hamilton–Susmel model on the weekly NYSE returns

Parameter	$\hat{\theta}$	$\hat{\sigma}$	Median	5% quantile	95% quantile	AR[a]
c	0.3307	0.0847	0.3278	0.2010	0.4883	0.3824
ϕ_1	0.3388	0.1429	0.3481	0.0917	0.5756	0.3824
ψ_1	−0.0930	0.1512	−0.0970	−0.3374	0.1776	0.3824
w	2.1884	1.0259	2.0585	0.7517	4.2162	0.4107
α	0.1785	0.0347	0.1766	0.1253	0.2421	0.3992
β	0.4119	0.0988	0.4165	0.2384	0.5745	0.3918
g_1	0.1477	0.0877	0.1231	0.0620	0.3506	–
g_2	0.5672	0.3466	0.4678	0.2372	1.4084	–
g_3	1.1706	0.7623	0.9489	0.4375	2.9625	–
v	6.3998	1.0703	6.0000	5.0000	8.0000	–
$\pi_{1,1}$	0.9850	0.0108	0.9893	0.9689	0.9981	–
$\pi_{1,2}$	0.0094	0.0097	0.0068	0.0000	0.0201	–
$\pi_{1,3}$	0.0056	0.0062	0.0019	0.0000	0.0105	–
$\pi_{2,1}$	0.0029	0.0020	0.0031	0.0000	0.0138	–
$\pi_{2,2}$	0.9861	0.0094	0.9888	0.9669	0.9967	–
$\pi_{2,3}$	0.0110	0.0082	0.0068	0.0002	0.0243	–
$\pi_{3,1}$	0.0115	0.0092	0.0113	0.0000	0.0267	–
$\pi_{3,2}$	0.0172	0.0106	0.0155	0.0020	0.0368	–
$\pi_{3,3}$	0.9713	0.0207	0.9774	0.9288	0.9932	–

Note: [a]Acceptance rate.

Table 2. Estimated parameters and posterior statistics of our model on the weekly returns from the NYSE

Parameter	$\hat{\theta}$	$\hat{\sigma}$	Median	5% quantile	95% quantile	AR[a]	$\hat{\theta}$[b]
c	0.3168	0.0788	0.3115	0.1977	0.4531	0.3211	0.2890
ϕ_1	0.3642	0.1345	0.3727	0.1254	0.5699	0.3211	0.4414
ψ_1	−0.1208	0.1438	−0.1270	−0.3519	0.1266	0.3211	−0.2058
w_1	0.3298	0.0694	0.3216	0.1765	0.5116	0.5374	0.2619
α_1	0.3517	0.1369	0.3456	0.1455	0.5858	0.5417	0.3510
β_1	0.1780	0.1182	0.1576	0.0225	0.3979	0.5345	0.2687
w_2	1.1423	0.3944	1.1304	0.5217	1.7951	0.4243	1.1237
α_2	0.1082	0.1333	0.1029	0.0314	0.2032	0.4201	0.0979
β_2	0.3794	0.1840	0.3759	0.0846	0.6731	0.4192	0.3745
w_3	3.9438	1.1507	3.9218	2.0629	5.9045	0.4932	3.8605
α_3	0.1804	0.0521	0.1730	0.0759	0.2978	0.4888	0.1559
β_3	0.2658	0.1467	0.2535	0.0466	0.5288	0.4923	0.2562
v	7.5089	1.5672	7.0000	5.0000	10.0000	–	6.7255
$\pi_{1,1}$	0.9870	0.0098	0.9893	0.9679	0.9981	–	0.9928
$\pi_{1,2}$	0.0090	0.0082	0.0068	0.0004	0.0254	–	0.0072
$\pi_{1,3}$	0.0040	0.0062	0.0013	0.0000	0.0167	–	0.0000
$\pi_{2,1}$	0.0010	0.0015	0.0012	0.0000	0.0040	–	0.0000
$\pi_{2,2}$	0.9818	0.0111	0.9835	0.9609	0.9968	–	0.9817
$\pi_{2,3}$	0.0172	0.0106	0.0155	0.0028	0.0371	–	0.0183
$\pi_{3,1}$	0.0033	0.0038	0.0019	0.0000	0.0112	–	0.0024
$\pi_{3,2}$	0.0319	0.0178	0.0301	0.0060	0.0646	–	0.0248
$\pi_{3,3}$	0.9648	0.0175	0.9668	0.9329	0.9898	–	0.9709

Notes: [a]Acceptance rate.
[b]Parameter estimates for a model specification as in Haas *et al*. (2004).

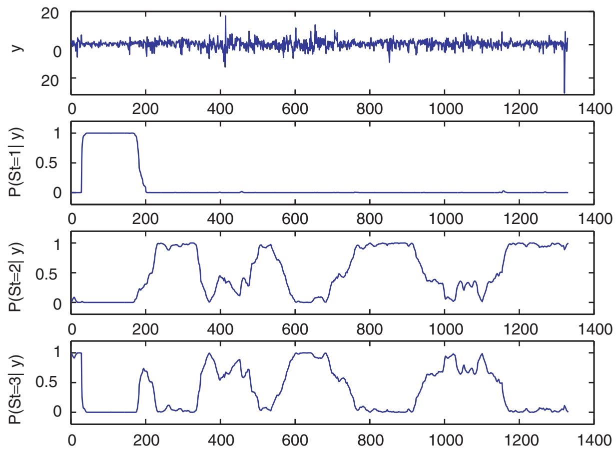

Fig. 1. Posterior probabilities of the different regimes for the modified Hamilton–Susmel model on the weekly NYSE returns. The corresponding parameter estimates are found in Table 1

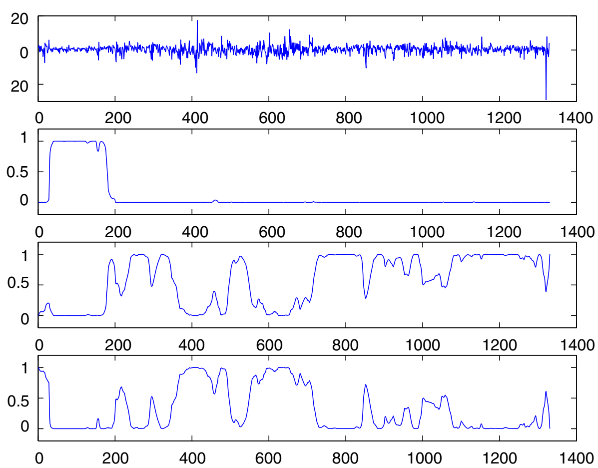

Fig. 2. Posterior probabilities of the regime 1–3 for our model. The corresponding parameter estimates are found in Table 2

parameter space of our model is of higher dimension. We therefore conclude that full MS–ARMA–GARCH models outperform models such as the Hamilton–Susmel model.

Comparison to the model specification due to Haas et al. (2004)

We estimated the parameters from an equivalent model specification as given by Haas *et al.* (2004) through an Expectation–Maximization (EM) algorithm and report these in the last column of Table 2. The results are virtually the same as our MCMC

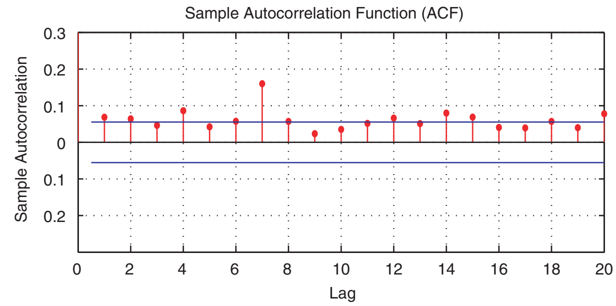

Fig. 3. Estimated conditional posterior autocorrelation of the squared residuals computed without the crash in 1987 for the modified Hamilton–Susmel model parameterized with the estimates from Table 1
Note: The bands are the approximate confidence bounds for a white noise process.

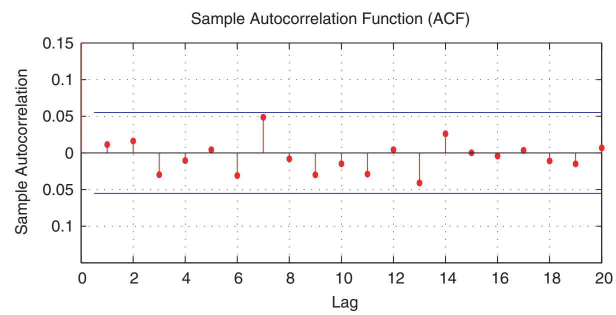

Fig. 4. Estimated conditional posterior autocorrelations of the squared residuals computed without the crash in 1987 for our model parameterized with the estimates from Table 2
Note: The bands are the approximate confidence bounds for a white noise process.

estimates and the diagnostics indicate a similar goodness-of-fit (Table 4).

To further compare these two models and estimators we examine the US dollar/yen foreign exchange rate.[10] Based on the results reported in Haas *et al.* (2004) we expect that the models should differ more significantly for this dataset since the regimes are estimated to switch more frequently. Furthermore, we conclude that the model fit deteriorates if one introduces a third regime. Hence, we compute the estimates of our and their model specification for a two regime GARCH(1,1) model.

With Gaussian innovations the model fit is rather poor and we only report the estimates for Student *t*-distributed innovations in Table 3 and the in-sample diagnostics in Table 4. We see that the transition probabilities for the second regime are significantly different. The impact of this becomes apparent if we examine the posterior probabilities of the second

[10] We evaluate the estimator on the percentage log returns of the US dollar/yen from 2 January 1990 to 3 October 2003, $T = 3460$.

Table 3. Estimated parameters and posterior statistics on the log returns of the US dollor/yen foreign exchange rate

Parameter	$\hat{\theta}$	$\hat{\sigma}$	Acceptance rate	$\hat{\theta}$[a]
w_1	0.0049	(0.0011)	0.21	0.0031
α_1	0.0120	(0.0041)	0.19	0.0151
β_1	0.9618	(0.0086)	0.22	0.9625
w_2	0.2706	(0.1490)	0.26	0.0101
α_2	0.1542	(0.0591)	0.24	0.0651
β_2	0.4818	(0.1561)	0.25	0.9565
v	6.0248	(0.6087)	–	6.670
$\pi_{1,1}$	0.9960	(0.0013)	–	0.9739
$\pi_{2,2}$	0.9690	(0.1320)	–	0.6649

Notes: [a]As in Haas *et al*. (2004).

Table 4. Diagnostics on the estimated transformed residuals $\hat{u}_t$

				US dollar/yen FX			
Data model	$\{\epsilon_t\}$	NYSE MS (3)[a]	MS (3)[b]	$\{\epsilon_t\}$	MS (2)[a]	MS (2)[b]	CV[c] (5%)
Mean	0.3333	−0.0616	−0.0525	−0.0007	−0.0065	−0.0168	–
StdDev	2.3568	1.1404	0.9559	0.0071	1.0133	0.9587	–
Skewness	−0.1006	−0.0894	−0.0976	−0.5516	−0.1522	−0.100	–
Kurtosis	7.5284	2.7141	2.8184	7.1741	3.0156	2.5679	–
LM_{ARCH} (1)	–	0.8392	0.6095	–	0.8610	3.7101	3.8415
LM_{ARCH} (5)	–	3.3811	2.0758	–	5.2790	10.387	11.070
LM_{ARCH} (10)	–	13.569	10.641	–	7.0465	12.486	18.307
Jarque–Bera	–	5.924	7.193	–	12.912	32.909	5.9815

Notes: [a]Our model formulation.
[b]Model formulation as in Haas *et al*. (2004).
[c]Critical value.

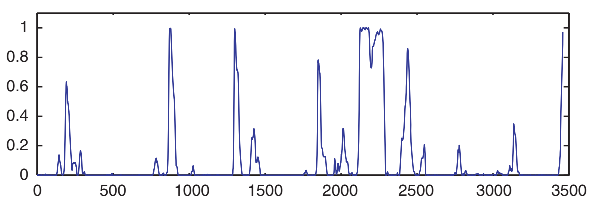

Fig. 5. Posterior probabilities of the second regime for our MS–GARCH formulation parameterized with the estimates in Table 3

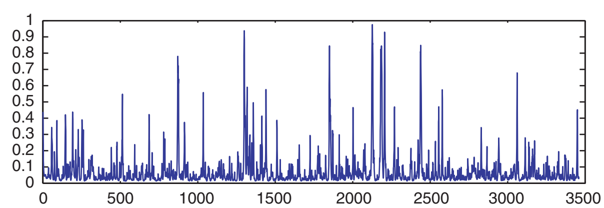

Fig. 6. Posterior probabilities of the second regime for the Haas *et al.* (2004) MS–GARCH formulation parameterized with the estimates in Table 3

regime depicted in Figs 5 and 6. In the Haas *et al.* (2004) model specification, the second regime is employed to capture parts of the excess kurtosis and skewness which is not accounted for in the innovations and through the GARCH parameters, whereas our formulation seems to capture more of a regime switching behaviour. However, both algorithms showed problems in the estimation which are characterized by a very slow convergence of the MCMC chain and a strong dependence on the starting parameters in the case of the EM-based algorithm.

V. Conclusions

In this article, we developed a MCMC method to compute the parameter estimates for a full MS–ARMA–GARCH model. These models are regarded as a promising class to describe certain phenomena in econometric time series observed in different markets. They seem appealing since they can tell a 'story' that is easily interpreted. But due to their severe intractability, their predictive power is difficult to assess in practice. In fact, the only thing that is straightforward about MS models is their specification. Due to their full path dependence, models with moving average or GARCH components could not be estimated. But this does not allow one to unleash the full power of ARMA–GARCH models. The algorithm that we presented overcomes this problem and is an important step for the further study of MS models. It can be easily extended and employed for all specifications of MS models.

If the data show no signs of switching in the mean, the Haas–Mittnik–Paolella model seems to be a good alternative to our specification, since it is much less intensive to estimate and makes an out-of-sample goodness of fit assessment feasible.

A lot of attention is directed towards these models at the moment and other research, such as that of Francq and Zakoian (2001), has established important results regarding the conditions for stationarity or the tail behaviour of these models. Though MS models are hard to handle, considerable progress has been made and we are confident that with increasing computational power these models are a valuable tool for financial modellers.

Acknowledgements

S.T. Rachev gratefully acknowledges research support by grants from the Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the German Research Foundation (DFG) and the German Academic Exchange Service (DAAD).

References

Bauwens, L. and Lubrano, M. (1998) Bayesian inference on GARCH models using the Gibbs sampler, *Econometrics Journal*, **1**, C23–46.

Cai, J. (1994) A Markov model of switching-regime ARCH, *Journal of Business and Economic Statistics*, **12**, 309–16.

Carlin, B. P., Polson, N. G. and Stoffer, D. S. (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modeling, *Journal of the American Statistical Association*, **87**, 493–500.

Chib, S. and Greenberg, E. (1994) Bayes inference in regression models with ARMA(p,q) errors, *Journal of Econometrics*, **64**, 183–206.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1008.

Engle, R. F. and Bollerslev, T. (1986) Modelling the persistence of conditional variances, *Econometric Reviews*, **5**, 1–50.

Eraker, B., Jacquier, E. and Polson, N. G. (1998) Pitfalls in MCMC algorithms, Technical Report, Department of Econometrics and Statistics, Graduate School of Business, University of Chicago, pp. 98–104.

Francq, C. and Zakoian, J. M. (2001) Stationarity of multivariate Markov-Switching ARMA models, *Journal of Econometrics*, **102**, 339–64.

Francq, C. and Zakoian, J. M. (2002) L^2-structures of standard and switching-regime GARCH models and their implications for statistical inference, Working Paper, University du Littoral-Cote d'Opale, Universite de Lille 3 and CREST.

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

Haas, M., Mittnik, S. and Paolella, M. S. (2004) A new approach to Markov-Switching GARCH models, *Journal of Financial Econometrics*, **2**, 493–530.

Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357–84.

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

Henneke, J. S., Rachev, S. T. and Fabozzi, F. (2006) MCMC methods for the estimation of MS–ARMA–GARCH Models, Technical Report, Universität Karlsruhe.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2004) Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics*, **122**, 185–212.

Kaufmann, S. and Fruewirth-Schnatter, S. (2000) Bayesian analysis of switching ARCH models, Working Paper, University of Vienna.

Lehmann, E. (1975) *Nonparametrics: Statistical Methods Based on Ranks*, McGraw-Hill, New York.

Maekawa, K., Lee, S., and Tokutsu, Y. (2005) A note on volatility persistence and structural changes in GARCH models, Technical Report, University of Hiroshima.

Nakatsuma, T. (1998) A Markov-chain sampling algorithm for GARCH models, *Studies in Nonlinear Dynamics and Econometrics*, **3**, 107–17.

Rapach, D. E. and Strauss, J. K. (2005) Structural breaks and GARCH models of exchange rate volatility, Technical Report, Saint Louis University.

Robert, C. P. (1994) *The Bayesian Choice*, Springer-Verlag, New York.

Robert, C. P. and Casella, G. (1999) *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, New York.

Smith, D. R. (2002) Markov-Switching and stochastic volatility diffusion models of short-term interest rates, *Journal of Business and Economic Statistics*, **20**, 183–97.

Tanner, M. A. and Wong, W. H. (1987) The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528–49.

Tierny, L. (1994) Markov chains for exploring posterior distributions, *Annals of Statistics*, **22**, 1701–62.

Appendix A: Model Specification of Haas *et al*. (2004)

Let the univariate time series ε_t be given by

$$\varepsilon_t = \sqrt{h_t(S_t)} \cdot u_t$$
$$u_t \sim N(0, 1)$$

where the conditional variance h_t is an $(\mathcal{S} \times 1)$-dimensional vector process given by

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1}$$

with

$$\alpha_i = (\alpha_{i,1}\ \alpha_{i,2} \cdots \alpha_{i,\mathcal{S}})'$$
$$\beta = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_{\mathcal{S}})$$

where $h_t(S_t)$ denotes the element of h_t at position S_t.

Appendix B: Stationarity Conditions of Francq and Zakoian (2002)

Define

$$\xi_{m',m''}(i) = \binom{m'}{m''} \{c(i)\}^{m'-m''} \frac{\mu_{2m'}}{\mu_{2m''}} \sum_{m'''}^{m''} \binom{m''}{m'''} \times \alpha(i) m''' \mu_{2m'''} \beta(i)^{m''-m'''}$$

where i identifies the regime and $\mu_{2m'}$ is the $2m'$-th moment of the innovation.

$$\mathbb{P}(f) = \begin{pmatrix} \pi_{1,1} f(1) & \ldots & \pi_{\mathcal{S},1} f(1) \\ \vdots & & \vdots \\ \pi_{1,\mathcal{S}} f(\mathcal{S}) & \cdots & \pi_{\mathcal{S},\mathcal{S}} f(\mathcal{S}) \end{pmatrix}$$

where $f : [1, \ldots, \mathcal{S}] \mapsto \mathbb{R}$

A MS–GARCH model as given by (1b) is $2m$-order stationary if $\mathbb{P}(\hat{o}(\xi_{2m,2m})) < 1$, where $\rho()$ denotes the biggest eigenvalue in modulus of a matrix.