

MARKOV SWITCHING IN STOCHASTIC VARIANCE. BAYESIAN COMPARISON OF TWO SIMPLE MODELS

ŁUKASZ KWIATKOWSKI¹

Katedra Metod Statystycznych
Krakowska Akademia
im. Andrzeja Frycza Modrzewskiego
PL 30-705 Kraków, ul. Gustawa Herlinga-Grudzińskiego 1
e-mail: kwiatkowski.lukas@gmail.com

Praca przedstawiona na posiedzeniu Komisji Nauk Ekonomicznych i Statystyki Oddziału PAN w Krakowie 13 stycznia 2009 roku przez autora.

ABSTRACT

Ł. Kwiatkowski. *Markov Switching in Stochastic Variance. Bayesian Comparison of Two Simple Models.* Folia Oeconomica Cracoviensia 2008–2009, 49–50: 109–143.

In the paper two particular Markov Switching Stochastic Volatility models (MSSV) are under consideration: one with a switching intercept in the log-volatility equation, and the other — with a regime-dependent autoregression parameter. While the former one is fairly common in the literature (as a tool of taking account for regimes of different mean volatility level), the latter has not been paid almost any attention so far. We note the fact, that state-varying mean volatility may arise from switches in the intercept or in the autoregression parameter. Hence, we aim to compare these two models in respect of goodness of fit to the data from the Polish financial market, employing Bayesian techniques of estimation and model comparison. Clear evidence of structural shifts in the volatility pattern is found. Two different regimes of the economy are characterized in terms of the mean volatility level and the variance of volatility.

KEY WORDS — SŁOWA KLUCZOWE

Markov switching, Stochastic Volatility model, Bayesian inference
przełączanie typu Markowa, model zmienności stochastycznej (SV), wnioskowanie bayesowskie

¹ Assistant at The Andrzej Frycz Modrzewski Krakow University College, Department of Statistics. The author is deeply indebted to Prof. J. Osiewalski and Dr. A. Pajor for their invaluable help and advice.

1. INTRODUCTION

Recent years have witnessed a rapid increase in the interest in modelling time-varying volatility of financial time series. Among the most popular tools devised to capture some of its common features have been two parametric model families: the ARCH processes, introduced by Engle (1982) (along with their numerous extensions starting with GARCH process of Bollerslev, 1986), and the stochastic volatility (SV) processes, of whose the main concept has been presented by Clark (1973). Although in formally different ways, in both of the above conditional variance equation is defined explicitly.

The underlying assumption of these constructions is homogeneity of the modelled time series, which means exclusion of potential structural breaks occurring in the analyzed period. It allows one to presume that the parameters of interest remain constant over time. However, volatility clustering, a common phenomenon observed in stock returns series, may question this belief. It is so due to some heuristic reasoning that less volatile periods alternating with these of higher uncertainty may somehow correspond with structural breaks occurring in the data. In view of potential heterogeneity of a certain time series, models such as GARCH or SV are of too restrictive nature (Hwang et al., 2004). Not being able to capture discrete shifts of states of the economy may be the cause for these models to yield somewhat misleading results. For instance, Granger and Hyung (1999) and Diebold and Inoue (2001) suggest that structural breaks in the mean of volatility may be a source of volatility persistence. It follows that a proper model should include an explicit mechanism capable of accounting for possible regime changes. One of the most popular in this respect is Markov switching (MS) mechanism introduced by Hamilton (1989). What he suggested is an autoregressive process whose parameters are subject to changes over time according to a latent homogeneous Markov chain. Since then many studies have been undertaken to employ the idea of MS into volatility models, mainly those of the GARCH family (see Bauwens et al., 2006, among many).

The aim of the paper is Bayesian estimation and comparison (in terms of goodness of fit to the data) three SV specifications: a non-switching basic stochastic volatility (BSV) model and two Markov Switching SV (MSSV) models (one with a regime-changing intercept and the other with a switching autoregression parameter in the volatility equation). The dataset comprises daily observations on the 1-month Warsaw Interbank Offered Rate (WIBOR1M) interest rates over the period from April 17, 2000 to April 7, 2008. Incidental to the analysis of the regime-switching constructions is a search for potential structural shifts occurring in the series and — if any are found — characterization of the identified states of the economy.

There are several reasons behind our research. Firstly, employing non-switching models in view of potential structural breaks in the time series may

lead to a model misspecification error. In this regard, switching specifications, like the MSSV processes, may be of value as they account for discrete shifts in the parameters. Secondly, we note that abrupt changes in the mean volatility level, which is the reason widely cited in the literature for employing the MSSV structures, may be attributed not only to the switching intercept but, alternatively, to the regime-changing autoregression parameter in the volatility equation. Finally, neither the issue above nor the MSSV models with a switching elasticity of volatility² are tackled in the literature known to the author.³

As regards the current state of the literature on the MSSV models, in a predominant part of the studies only two-state specifications with a switching intercept are concerned (Smith, 2000; Kalimipali and Susmel, 2001; Casarin, 2003; Shibata and Watanabe, 2005; Carvalho and Lopes, 2006). Three-state models are analyzed in So et al. (1998) and Hwang et al. (2003, 2004). In terms of the estimation tools the Bayesian approach prevails, with use of either standard MCMC procedures (the Gibbs sampler; So et al., 1998; Kalimipali and Susmel, 2001, Shibata and Watanabe, 2005) or more recent (auxiliary) particle filters (Casarin, 2003; Carvalho and Lopes, 2006). Some of the models feature additional elements such as term structure (Smith, 2000; Kalimipali and Susmel, 2001) and heavy-tailed distributions of the noise term in the observable process (Casarin, 2003).

We conduct the analysis in the Bayesian setting, which allows fully probabilistic inference on all the unknown quantities of the model as well as well-founded model comparison. As opposed to the 'classical' (i.e. non-Bayesian) tools, Bayesian methodology in the context of switching models (or, more generally, mixture models) is found even more appealing. The latter stems from the possibility of inference on the latent regimes unconditionally upon the parameter estimates (see Gärtner, 2007).

The remainder of the paper is organized as follows. In Section 2 we present the models under consideration and selected regime characteristics, of which use is made in the further parts. Bayesian estimation of the models and their comparison are briefly discussed in Section 3, followed by an empirical illustration of the presented methodology in Section 4. Finally, Section 5 concludes.

² 'Elasticity of volatility' is the term used by Smith (2000) with reference to the autoregression parameter, ϕ , in the log-volatility equation of a SV model given as: $\ln h_t = \mu + \phi \ln h_{t-1} + \sigma \eta_t$. Assuming $\eta_t \sim iid(0, 1)$ (i.e. each η_t is an independent and identically distributed random variable with zero mean and unit variance) the third parameter, σ , is a standard deviation of the innovation term $\sigma \eta_t$, and hence referred to as 'volatility of volatility'.

³ The only works in which the autoregression parameter is allowed to switch over the regimes are of Hwang et al. (2003, 2004). However, not only the estimation approach employed in these studies (Quasi-ML), but also the specification of the log-volatility equation is different than in our work.

2. SELECTED MARKOV SWITCHING SV (MSSV) MODELS

In this part the basics of selected MSSV processes are presented. We start with the following definition of a general Markov Switching SV process.

Definition 1

A stochastic process⁴ $\{y_t, t \in N \cup \{0\}\}$ follows a two-state Markov Switching Stochastic Volatility (MSSV) process if and only if for each $t \in N \cup \{0\}$ the following assumptions hold:

$$y_t = \varepsilon_t \sqrt{h_t}; \quad (1)$$

$$\ln h_t = \mu_{S_t} + \varphi_{S_t} \ln h_{t-1} + \sigma_{S_t} \eta_t; \quad (2)$$

$$\left\{ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \middle| t \in N \cup \{0\} \right\} \sim \text{iiN}^{(2)}(\mathbf{0}_{(2 \times 1)}, I_2);$$

$\{S_t, t \in N \cup \{0\}\}$ — a homogenous, ergodic and irreducible two-state Markov chain;

$$S_t \in Q = \{1, 2\};$$

$$\Pr(S_t = j \mid S_{t-1} = i) = p_{ij}, \quad p_{ij} \in (0, 1), \quad \sum_{j=1}^2 p_{ij} = 1, \quad i, j = 1, 2.$$

The observable variable, y_t , is defined as a product of a Gaussian white noise and conditional standard deviation.⁵ Equation 2 defines the log-volatility which evolves over time according to a simple switching autoregressive process. Since all of the parameters in the latter feature regime-changing property, the definition may be viewed quite general, although further extensions are possible (a heavy-tailed distribution for ε_t can be considered, for instance, as in Casarin, 2003). The switching mechanism, represented by the family of discrete random variables S_t 's, is assumed to follow a simple two-state Markov chain, in accordance with the idea proposed by Hamilton (1989). For the sake of our study, ergodicity and irreducibility of the chain are assumed by restricting the transition probabilities, p_{ij} , to lay strictly within the unit interval.

One should note, that a basic stochastic volatility process (BSV), with the log-volatility equation defined as:

$$\ln h_t = \mu + \varphi \ln h_{t-1} + \sigma \eta_t,$$

⁴ By N we denote the set of positive integers.

⁵ It is straightforward to show that — conditionally upon a σ -algebra with respect to which h_t is measurable — the latter constitutes conditional variance of the process $\{y_t\}$, i.e: $\text{Var}(y_t \mid F_{t-1}, \eta_t, S_t) = h_t$, where F_{t-1} is the past information about the process $\{h_t, t \in N \cup \{0\}\}$ up to the moment $t-1$.

may be viewed as a particular case of the general MSSV process once $\mu_1 = \mu_2$, $\varphi_1 = \varphi_2$ and $\sigma_1 = \sigma_2$ hold. However, the transition probabilities, p_{ij} , remain then unidentified.

In our work two special cases of the general MSSV process are of particular interest: the one with a switching intercept and the other — with a regime-dependent autoregression parameter. A concise discussion of both follows.⁶

2.1. MSSV model with a switching intercept, MSSV(μ)

In this case Equation 2 collapses to:

$$\ln h_t = \mu_{S_t} + \varphi \ln h_{t-1} + \sigma \eta_t = \begin{cases} \mu_1 + \varphi \ln h_{t-1} + \sigma \eta_t & \Leftrightarrow S_t = 1 \\ \mu_2 + \varphi \ln h_{t-1} + \sigma \eta_t & \Leftrightarrow S_t = 2 \end{cases}. \quad (3)$$

For the sake of identifiability of the model we reparametrize the switching parameter as (see So et al., 1998):

$$\mu_{S_t} = \gamma_1 + \gamma_2 I(S_t = 2),$$

where $\gamma_1 \in R$, $\gamma_2 < 0$ and $I(\cdot)$ denotes the indicator function which takes one if the condition in the parentheses is satisfied and zero otherwise. Such a representation of the switching intercept results in inequality $\mu_1 > \mu_2$. It may be shown that the latter is equivalent to predetermining states '1' and '2' as ones of high and low mean log-volatility level, respectively, that is:

$$\mu_1 > \mu_2 \Leftrightarrow E(\ln h_t | S_t = 1) > E(\ln h_t | S_t = 2).$$

For the model in question we shall also assume covariance stationarity of the log-volatility process following Equation 3, for which it is necessary and sufficient⁷ to guarantee that $|\varphi| < 1$.

2.2. MSSV model with a switching autoregression parameter, MSSV(φ)

In our study we note that discrete shifts in the mean volatility level may result from not only a switching intercept, but — alternatively — a regime-changing autoregression parameter. Hence, we consider a MSSV model with Equation 2 assuming the form:

$$\ln h_t = \mu + \varphi_{S_t} \ln h_{t-1} + \sigma \eta_t, \quad (4)$$

⁶ For a comprehensive work on non-switching SV models we refer to Pajor (2003).

⁷ General results on second-order and strict stationarity of switching vector autoregression processes may be found in Francq and Zakoian (2001).

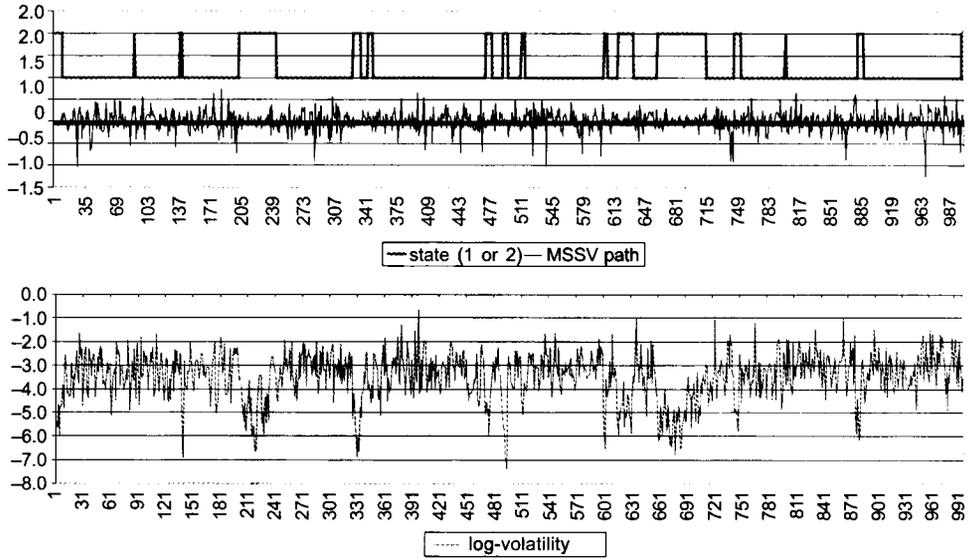


Fig. 1. Simulated path of a MSSV(φ) process ($\mu = -2.5$; $\varphi_1 = 0.2$; $\varphi_2 = 0.5$; $\sigma^2 = 0.6132$; $p_{11} = 0.98$; $p_{22} = 0.95$) and the corresponding log-volatility and regime-switching processes

For the identifiability reasons we shall impose the inequality $\varphi_1 < \varphi_2$. Once formulas for conditional expectations $E(\ln h_t | S_t = i)$ ($i = 1, 2$) are available (see the following subsection), it is easily shown that:

$$\mu < 0 \Rightarrow [\varphi_1 < \varphi_2 \Leftrightarrow E(\ln h_t | S_t = 1) > E(\ln h_t | S_t = 2)]$$

and

$$\mu > 0 \Rightarrow [\varphi_1 < \varphi_2 \Leftrightarrow E(\ln h_t | S_t = 1) < E(\ln h_t | S_t = 2)],$$

which indicates different mean volatility levels in each of the regimes.

Further, we assume covariance stationarity of the log-volatility process following Equation 4. The relevant (necessary and sufficient) condition is given by the set of inequalities:⁸

$$\begin{cases} R_1 < 1 \\ R_2 < 2, \end{cases}$$

where:

$$R_1 = p_{11}\varphi_1^2 + p_{22}\varphi_2^2 + (1 - p_{11} - p_{22})\varphi_1^2\varphi_2^2,$$

$$R_2 = p_{11}\varphi_1^2 + p_{22}\varphi_2^2.$$

⁸ The condition is also valid for the general case, in which all the three volatility parameters are allowed regime shifts (see Francq and Zakoian, 2001). One should note that the condition is somewhat contrary to an initial 'intuition' according to which it should be 'enough' to assume that $|\varphi_1| < 1$ and $|\varphi_2| < 1$. The latter constitutes neither a necessary nor a sufficient condition for second-order stationarity of a two-state switching first-order autoregression (ibid.).

Figure 1 depicts a simulated path of a certain MSSV(φ) process and the corresponding regime-switching and (stationary) log-volatility processes. The latter displays evident shifts in the mean level (according to the switching mechanism), which manifest themselves in the form of volatility clustering.

2.3. Selected regime characteristics

While allowing different states of the economy, it is natural to characterize the regimes in some systematic way. In our work we do so by calculating selected regime-specific characteristics both of the log-volatility process and the switching mechanism as well, including:

— state-conditional mean log-volatility level:⁹

$$E_1 \equiv E(\ln h_t | S_t = 1) = \frac{\mu_1(1 - \varphi_2 p_{22}) + \mu_2 \varphi_1(1 - p_{11})}{1 - \varphi_1 \varphi_2 - \varphi_2 p_{22}(1 - \varphi_1) - \varphi_1 p_{11}(1 - \varphi_2)},$$

$$E_2 \equiv E(\ln h_t | S_t = 2) = \frac{\mu_2(1 - \varphi_1 p_{11}) + \mu_1 \varphi_2(1 - p_{22})}{1 - \varphi_1 \varphi_2 - \varphi_2 p_{22}(1 - \varphi_1) - \varphi_1 p_{11}(1 - \varphi_2)};$$

— state-conditional variance of the log-volatility process:

$$V_i \equiv \text{Var}(\ln h_t | S_t = i) = E(\ln^2 h_t | S_t = i) - E_i^2 \quad \text{for } i = 1, 2,$$

where:

$$E(\ln^2 h_t | S_t = 1) = \frac{d_1(1 - \varphi_2^2 p_{22}) + d_2 \varphi_1^2(1 - p_{11})}{1 - \varphi_1^2 p_{11} - \varphi_2^2 p_{22} + \varphi_1^2 \varphi_2^2(-1 + p_{11} + p_{22})},$$

$$E(\ln^2 h_t | S_t = 2) = \frac{d_2(1 - \varphi_1^2 p_{11}) + d_1 \varphi_2^2(1 - p_{22})}{1 - \varphi_1^2 p_{11} - \varphi_2^2 p_{22} + \varphi_1^2 \varphi_2^2(-1 + p_{11} + p_{22})},$$

and

$$d_i = \mu_i^2 + 2\mu_i \varphi_i E(\ln h_{t-1} | S_t = i) + \sigma_i^2 \quad \text{for } i = 1, 2,$$

$$E(\ln h_{t-1} | S_t = i) = \sum_{j=1}^2 p_{ji}^* E(\ln h_t | S_t = j)$$

with $p_{ji}^* = \text{Pr}(S_{t-1} = i | S_t = j)$ being the inverse transition probabilities;¹⁰

⁹ First- and second-order state-conditional moments of the log-volatility process have been obtained for the general case (that is the one in which all three parameters are regime-changing) under assumption of covariance stationarity of that process and based on the results of Nielsen and Olesen (2000).

¹⁰ In the case of a two-state Markov chain the inverse transition probabilities defined as $p_{ji}^* = \text{Pr}(S_{t-1} = i | S_t = j)$ are easily shown to equal the ordinary transition probabilities, $p_{ji} = \text{Pr}(S_t = i | S_{t-1} = j)$.

— ergodic probabilities:¹¹

$$\pi_1 = \Pr(S_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}},$$

$$\pi_2 = \Pr(S_t = 2) = 1 - \pi_1;$$

— expected duration¹² of state 'i' (once the system has switched to that state; see Hamilton, 1989):

$$Dur_i = \frac{1}{1 - p_{ii}} \quad \text{for } i = 1, 2.$$

3. BAYESIAN ESTIMATION AND COMPARISON OF THE MSSV MODELS¹³

Estimation of the MSSV models is not trivial. Handling the maximum likelihood procedure is riddled with serious numerical obstacles due to the presence of (as much as) two latent processes underlying the observable process: the conditional volatilities, h_t 's, and the states, S_t 's. In our work we resort to Bayesian methodology, which prevails in the MSSV literature.¹⁴ Although new methods — based on the (auxiliary) particle filters — have been developed of recent,¹⁵ we employ the already 'classical' MCMC procedures: the Gibbs sampler and the Metropolis-Hastings algorithm, to simulate from the joint posterior distribution of all the unknown quantities of the model.

Let $y = (y_1, y_2, \dots, y_T)'$ denote the modelled time series, vector $h = (h_1, h_2, \dots, h_T)'$ be the series of the latent conditional volatilities and vector $S = (S_1, S_2, \dots, S_T)'$ — the unobserved Markov chain. We define the parameter vector as $\theta = (\beta', \sigma^2, p_{11}, p_{22})'$ with σ^2 , p_{11} and p_{22} being the parameters common to both MSSV models, and β comprising the model-specific parameters:

$$\beta = \begin{cases} (\gamma_1, \gamma_2, \varphi)' & \text{for } MSSV(\mu) \\ (\mu, \varphi_1, \varphi_2)' & \text{for } MSSV(\varphi) \end{cases}$$

Further, we employ the data-augmentation technique introduced by Tanner and Wong (1987), within which all the unknown quantities of the model

¹¹ Ergodic probabilities calculated for an ergodic Markov chain tell us approximately for how long (in terms of a part of the analyzed time series) the chain remains in each of its states. In the study we assume that $p_{ii} \in (0,1)$ for $i = 1, 2$, which ensures ergodicity of the switching process.

¹² Expected duration of a certain regime is calculated conditionally upon being in that state.

¹³ In the paper we discuss only the estimation of the switching SV models. For a detailed description of the Bayesian estimation of simple SV constructions we refer to Pajor (2003).

¹⁴ For the relevant references see Section 1.

¹⁵ Ditto.

are treated as random variables subject to estimation and taking values in the common space:

$$\omega' = (\theta', h', S') \in \Omega = \Theta \times H \times Q^T,$$

where $\theta \in \Theta \subset R^6$, $h \in H \subseteq R_+^T$, $S \in Q^T$ and $Q = \{1, 2\}$.

The joint posterior distribution of ω is factorized as:¹⁶

$$p(\theta, h, S | y) \propto p(y | h) p(h | S, \theta) p(S | \theta) p(\theta), \quad (5)$$

which reveals its hierarchical structure. Individual components of (5) are presented in the Appendix. Here, we focus on the prior structure of the parameters, for (almost) all of which mutual independence is assumed:

$$p(\theta) = \begin{cases} p(\beta) p(\sigma^2) p(p_{11}) p(p_{22}) & \text{for MSSV}(\mu), \\ p(\beta | p_{11}, p_{22}) p(\sigma^2) p(p_{11}) p(p_{22}) & \text{for MSSV}(\varphi). \end{cases}$$

Conditioning on the transition probabilities in $p(\beta | p_{11}, p_{22})$ for the model with a switching autoregression parameter stems from imposing prior restrictions of covariance stationarity of the underlying log-volatility process (see Section 2.2).

In the study we choose fairly diffuse priors, letting the posterior results arise mainly from the information contained in the data. More specifically, we have:

1. prior distributions for the parameters common to both MSSV models:¹⁷

$$- p(\sigma^2) = f_{IG}(\sigma^2 | v_1, v_2), v_1 = 1, v_2 = 200;$$

$$- p(p_{ii}) = f_B(p_{ii} | a_i, b_i), a_i = b_i = 1, \text{ for } i = 1, 2;$$

where $f_B(p_{ii} | a_i, b_i)$ denotes the density function of a Beta-distributed random variable, p_{ii} , with the shape and scale parameters equal a_i and b_i , respectively;

2. prior distribution for β :

— for the MSSV(μ) model:

$$p(\beta) = f_N^{(3)}(\beta | \beta_0, A_0^{-1}) I(\gamma_2 < 0) I(|\varphi| < 1), \beta_0 = 0_{(3 \times 1)}, A_0 = 0,01 \cdot I_3,$$

where $f_N^{(k)}(\beta | \beta_0, A_0^{-1})$ denotes the density function of a normally distributed k -variate random variable, β , with the mean vector and the precision matrix equal β_0 and A_0 , respectively;

¹⁶ The analysis is conducted conditionally on $h_0 = 1$, dependence on which is omitted in the notation.

¹⁷ We parametrize the density of the inverse gamma distribution as:

$$p(\sigma^2) = f_{IG}(\sigma^2 | v_1, v_2) = \frac{v_2^{-v_1}}{\Gamma(v_1)} \frac{1}{(\sigma^2)^{v_1+1}} \exp\left(-\frac{1}{v_2 \sigma^2}\right).$$

— for the MSSV(φ) model:

$$p(\beta | p_{11}, p_{22}) = f_N^{(3)}(\beta | \beta_0, A_0^{-1})I(R_1 < 0)I(R_2 < 0), \beta_0 = 0_{(3 \times 1)}, A_0 = 0,01 \cdot I_3.$$

As regards prior distributions for the parameters of the BSV model, we follow the structure employed in Pajor (2003), namely:

$$\begin{aligned} &— p(\sigma^2) = f_{IG}(\sigma^2 | v_1, v_2), v_1 = 1, v_2 = 200; \\ &— p(\mu, \varphi) = f_N^{(2)}(\mu, \varphi | 0_{(2 \times 1)}, A_0^{-1})I(|\varphi| < 1), A_0 = 0,01 \cdot I_3. \end{aligned}$$

The prior structure presented above provides very convenient (in terms of the sampling method) conditional posterior distributions of the model parameters.¹⁸ The latter are employed to construct a hybrid chain within the MCMC procedure, through which a N -sized sample from the joint posterior distribution is obtained, $\{\omega^{(q)}\}_{q=M+1}^{M+N}$, where q denotes the number of the cycle of the sampling algorithm, of which the first M cycles are discarded, and $\omega^{(q)}$ signifies the outcome on ω from the q -th step. Once the algorithm is complete, it is straightforward to obtain also a sample of any measurable function of ω , such as regime characteristics, in particular.

In order to allow Bayesian model comparison, the marginal likelihood for each of the estimated models needs to be evaluated. In our work we resort to the procedure introduced by Newton and Raftery (1994), in which the quantity of interest is estimated as:

$$\hat{p}(y | M_i) = \left[\frac{1}{N} \sum_{q=M+1}^{M+N} \frac{1}{p(y | \omega_{(i)}^{(q)}, M_i)} \right]^{-1},$$

where $\hat{p}(y | M_i)$ is the estimator of the marginal likelihood in the i -th model, M_i . Despite its lamentable numerical instability, the method proved satisfactory in our applications. Finally, to compare the models pair-wise use is made of Bayes factors, B_{ij} , defined as:

$$B_{ij} = \frac{p(y | M_i)}{p(y | M_j)}.$$

¹⁸ Full details on the posterior structure of all the estimated quantities are found in the Appendix.

4. EMPIRICAL STUDY

4.1. Data description

The methodology presented above is illustrated with an empirical study in which data from the Polish financial market is analyzed. More specifically, we consider a series of daily quotations of the 1-month Warsaw Interbank Offered Rate (WIBOR1M) interest rates over the period from April 17, 2000 to April 7, 2008 (which makes the total of 2002 observations). The series is plotted in Figure 2.

WIBOR1M interest rates (2000.04.17–2008.04.07)

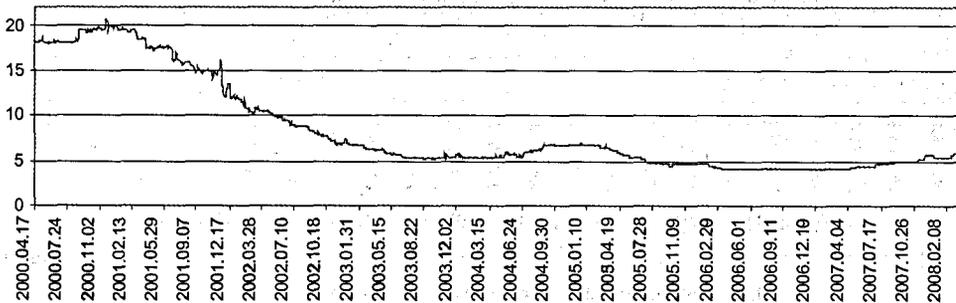


Fig. 2. The series of WIBOR1M interest rates, w_t

We calculate the daily log-returns, r_t 's, on the WIBOR1M interest rates, defined as:

$$r_t = 100 \ln(w_t / w_{t-1}),$$

where w_t denotes the price of the instrument at time t . The series of r_t 's is presented in Figure 3.

Log-returns on WILBOR1M (2000.04.18–2008.04.07)

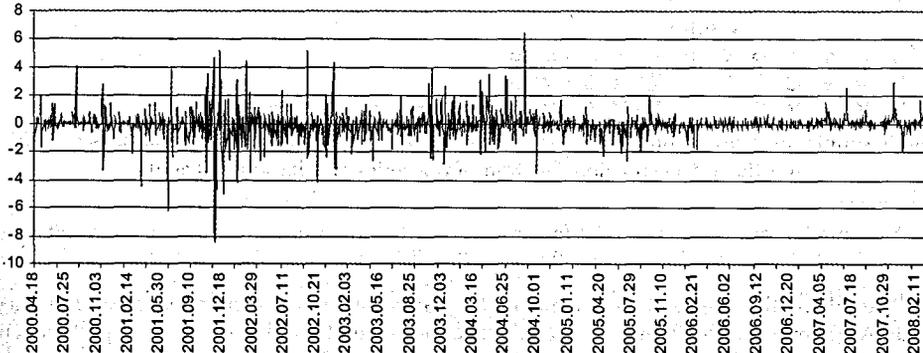


Fig. 3. Daily log-returns, r_t , on the WIBOR1M (April 18, 2000–April 7, 2008)

Further, we apply a simple linear filter — a first-order autoregressive model — to the data as to account for possibly non-zero conditional mean of the data generating process.¹⁹ Henceforth, the analysis is conducted for the resulting series of AR(1)-residuals²⁰ (see Fig. 4), denoted as y_t with $t = 1, 2, \dots, T = 2000$. The latter display features commonly found in financial data, including volatility clustering, high value of the empirical kurtosis coefficient (see Tab. 1), no

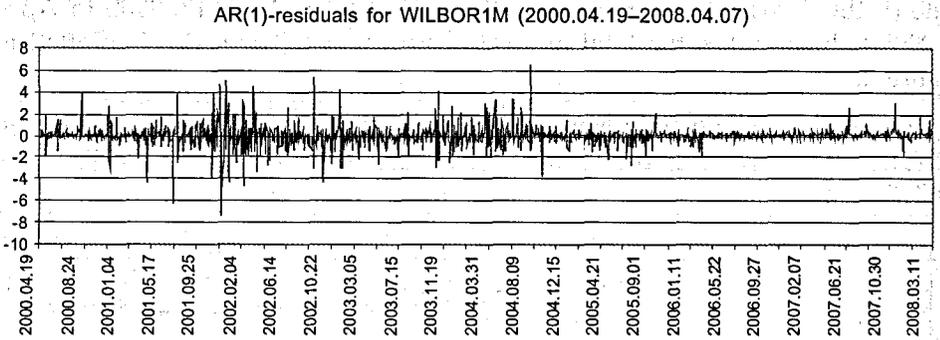


Fig. 4. The AR(1)-residuals for the daily log-returns on the WILBOR1M

Table 1

Descriptive statistics for AR(1)-residuals for WILBOR1M

Min	Max	Mean	Stand. deviation	Asymmetry	Kurtosis	ARCH(2) effect
-7.2629	6.5844	0.0000	0.8394	-0.4339	18.5818	$TR^2 = 177.1391$ ($p\text{-value} = 0.0000$)

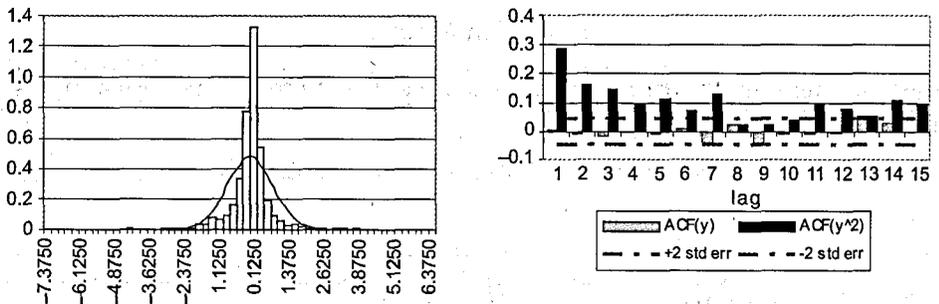


Fig. 5. Empirical distribution of the series $\{y_t, t = 1, 2, \dots, T\}$ with fitted normal density (left) and the autocorrelation function of the series and its square (right)

¹⁹ Estimation of the AR(1) model for the series of the log-returns, r_t 's, yields the results:

$$r_t = -0.0464 + 0.1537r_{t-1} + \hat{u}_t$$

(0.0188) (0.0221)

²⁰ An alternative approach is to simulate the parameters of the conditional mean modelled with an AR(1) process from their conditional posterior distributions. However, we expect that autocorrelations in the log-returns have little impact on the volatility and, hence, adopt the method used by So et al., 1998.

significant autocorrelations in the original series, yet strong auto-dependencies in the squared series (see Fig. 5). Additionally, left asymmetry in the empirical distribution of the residuals is found (see Tab. 1).

4.2. Results for the Basic SV model

For the estimation of the BSV model we employed the Gibbs sampler combined with the Metropolis-Hastings step for sampling the latent conditional volatilities, h_t 's, as done in Pajor (2003). The first $M = 500,000$ burnt-in iterations are discarded and the subsequent $N = 1,500,000$ iterations are regarded as a simulated sample from the joint posterior density.

Table 2 contains posterior means and standard deviations of the model parameters. One notes the posterior mean of the autoregression parameter, φ , being fairly close to unit. It is a common finding in the SV literature (see Pajor, 2003, among many), indicating evident persistence in the conditional volatility process. Despite prior independency between the parameters we observe strong posterior correlations (see Tab. 3). The latter may arise as a result of 'stabilization' of the unconditional characteristics of the volatility process, such as mean and variance.

Table 2

Posterior means and standard deviations (in parentheses) of the BSV parameters

μ	φ	σ^2
-0.3384 (0.0508)	0.8269 (0.0215)	1.1053 (0.1285)

Table 3

Posterior correlation matrix of the BSV parameters

M_1	μ	φ	σ^2
μ	1	0.8753	-0.6590
φ		1	-0.7185
σ^2			1

In Figure 6 the marginal prior and posterior distributions of the BSV parameters along with the plots of their ergodic means (against the number of cycles) are depicted. The results of posterior densities being of regular shapes and fast convergence of the ergodic means convergence to their posterior counterparts remain consistent with Pajor (2003).

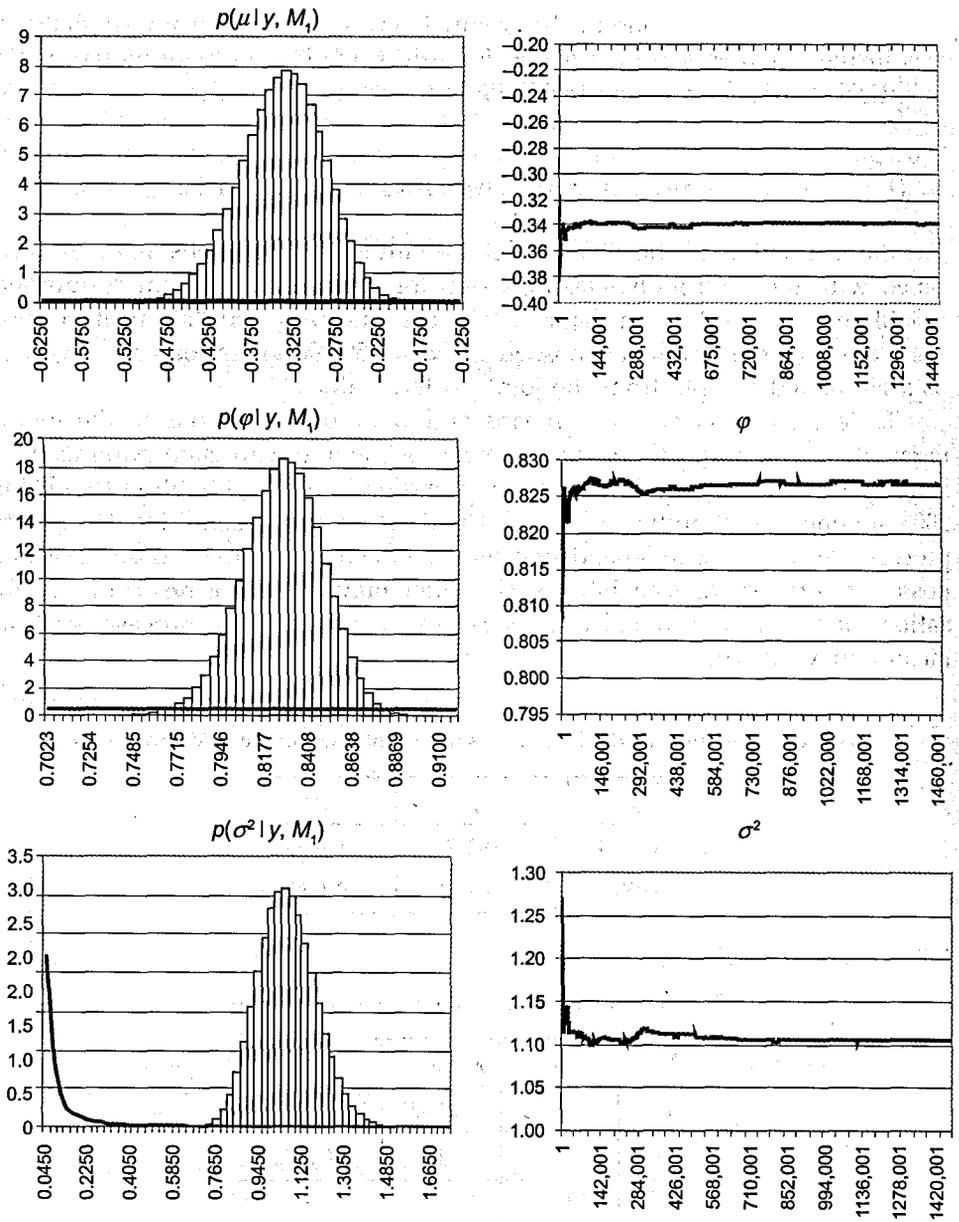


Fig. 6. Left column: marginal prior (solid line) and posterior distributions of the BSV parameters; right column: ergodic means of the parameters against the number of cycles

4.3. Results for the MSSV (μ) model

To estimate the model we employ the sampling algorithm presented in the Appendix.²¹ The first $M = 2,000,000$ burn-in iterations are discarded and the subsequent $N = 1,500,000$ iterations constitute a simulated sample from the joint posterior density.

As it can be gathered from the posterior means of the transition probabilities located very close to unit (see Tab. 4), the switching mechanism manifests strong persistence. Once a certain state is achieved, little is the probability of a switch to the alternative regime. Furthermore, one notes significantly negative posterior mean of γ_2 , which provides compelling evidence of discrete shifts in the value of the intercept. As compared with the results for the BSV model, the mean posterior of the elasticity of volatility is markedly lower. It is the most common finding cited in the MSSV literature, where it is argued that structural shifts unaccounted for by standard SV models may imply spuriously high persistence in the volatility process. However, we would not jump to such conclusions, unless the true autocorrelation functions of the log-volatility process in the BSV and MSSV model are surveyed.²² One may presume, that the very same 'spuriously' high volatility persistence implied by the non-switching SV specification may be captured by the switching counterpart, yet in a different manner (resulting, for instance, in the close-to-unit mean posterior probabilities p_{ii} , $i = 1, 2$). The issue merits further research.

Table 4

Posterior means and standard deviations of the parameters of model M_2

p_{11}	p_{22}	γ_1	γ_2	φ	σ^2
0.9960	0.9964	-0.2753	-0.7292	0.6658	1.3594
(0.0027)	(0.0026)	(0.0647)	(0.1051)	(0.0360)	(0.1437)

According to the posterior correlation matrix of the parameters (see Tab. 5), prior assumption of their mutual independence seems to be overruled by the data. In our belief, the non-zero posterior correlation coefficients may arise from 'stabilization' of the regime characteristics as well as the unconditional moments of the log-volatility process.

²¹ We note that the minimum acceptance rate while sampling h_t 's via the M-H algorithm, amounted to approximately 60%, which is found much satisfying.

²² For the purpose of comparison of volatility persistence implied by the BSV and both MSSV models, we averaged posterior empirical autocorrelation function (ACF) coefficients (lags: 1 to 15) for the sampled series of $\ln h_t$'s. The results appear not to reject the individual hypotheses of equal mean ACF coefficients across different models, therefore advocating the conjecture to follow in the main text.

Posterior correlation matrix of the parameters of model M_2

M_2	p_{11}	p_{22}	γ_1	γ_2	φ	σ^2
p_{11}	1	0.3966	-0.0062	0.1986	0.1292	0.0035
p_{22}		1	0.1524	0.2232	0.1523	0.0078
γ_1			1	0.1262	0.5273	-0.3921
γ_2				1	0.8012	-0.4871
φ					1	-0.6647
σ^2						1

Marginal posterior densities of the transition probabilities p_{ii} concentrate tightly on the left of unit (see Fig. 7), which indicates that the analyzed dataset is very informative with regard to the switching mechanism. Posterior distri-

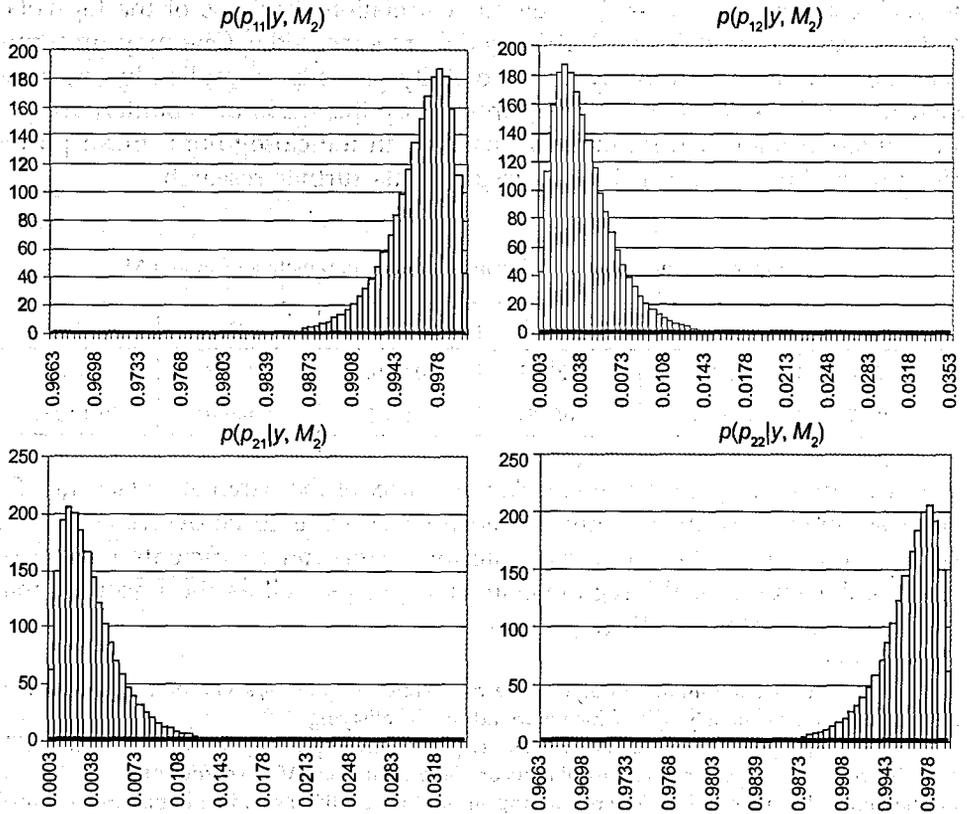


Fig. 7. Marginal prior (solid line) and posterior distributions of the transition probabilities in model M_2

butions of the remaining parameters are clearly unimodal and cluster (with slight asymmetries) around their means (see Fig. 8). Prior covariance stationarity of the log-volatility process is not rejected by the data, as the posterior density of the autoregression parameter, ϕ , clusters away on the left of unit.

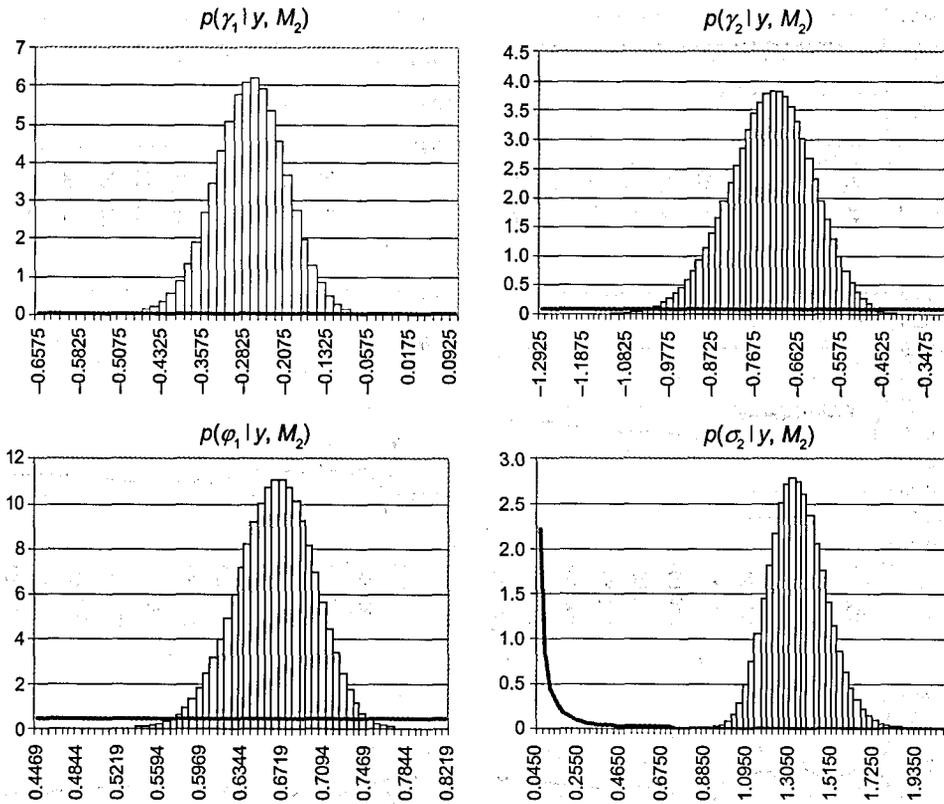


Fig. 8. Marginal prior (solid line) and posterior distributions of the log-volatility parameters in model M_2

A somewhat unstable behaviour of the ergodic means of the parameters (except for the transition probabilities) may raise concerns as regards convergence of the MCMC procedure (see Fig. 9). However, one should note rather negligible magnitude of the visible fluctuations.

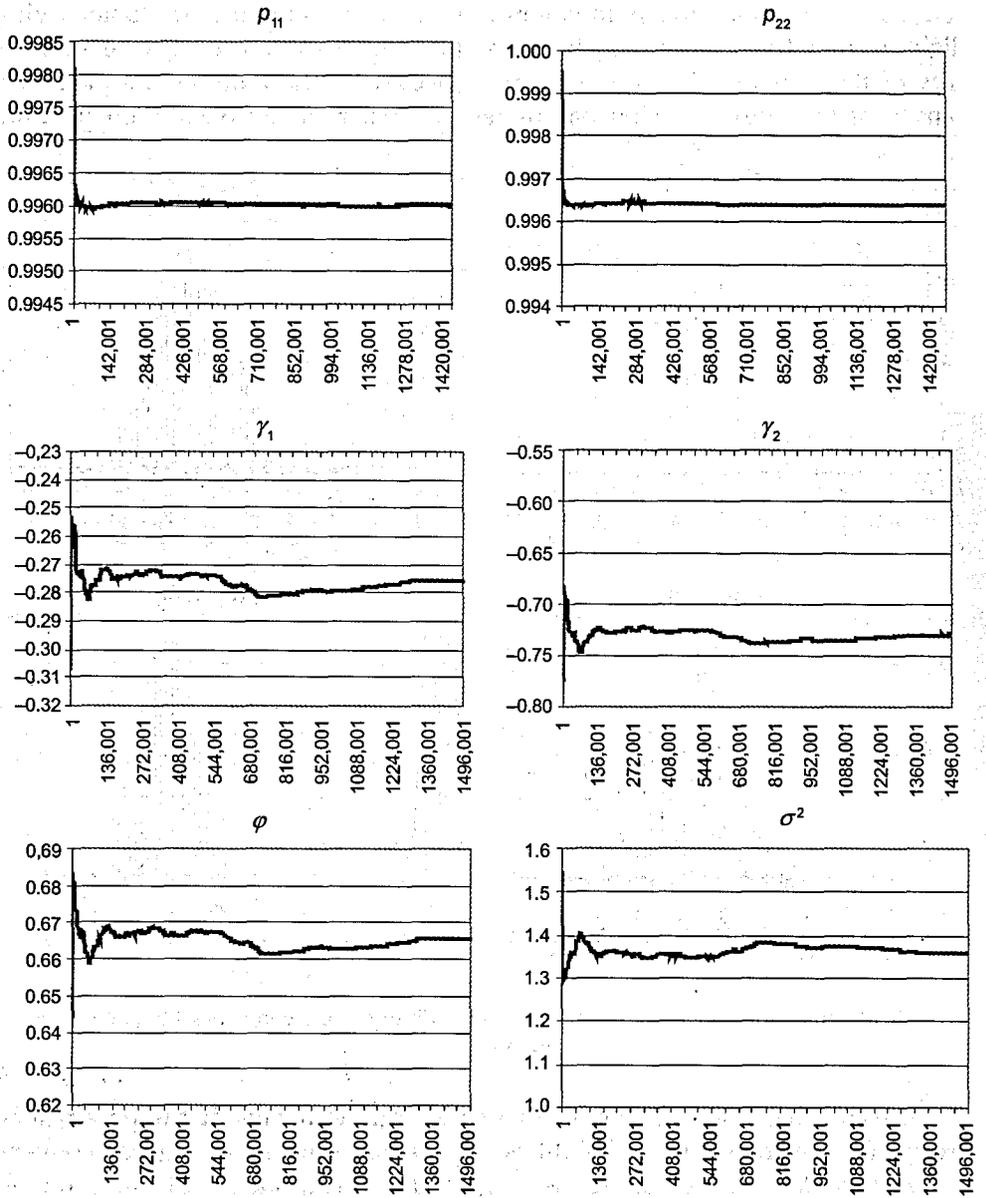


Fig. 9. Ergodic means of the parameters of model M_2 against the number of cycles

4.4. Results for the MSSV(φ) model

As in the previous case, the quantities of interest are sampled within the MCMC procedure presented in the Appendix.²³ The first $M = 2,000,000$ burn-in iterations are discarded and the subsequent $N = 1,500,000$ iterations constitute a simulated sample from the joint posterior density.

As far as the switching mechanism is concerned, similar (to the previous model) results are obtained. Posterior means of the probabilities p_{ii} are very

Table 6

Posterior means and standard deviations of the parameters of model M_3

p_{11}	p_{22}	μ	φ_1	φ_2	σ^2
0.9939	0.9961	-0.4511	0.5465	0.8201	1.2485
(0.0046)	(0.0026)	(0.0642)	(0.0869)	(0.0233)	(0.1375)

close to unit, implying high persistence in the latent Markov chain (see Tab. 6). Moreover, the posterior means of the switching parameter differ substantially across the two regimes. It follows that switches between two genuinely dis-

Table 7

Posterior correlation matrix of the parameters of model M_3

M_3	p_{11}	p_{22}	μ	φ_1	φ_2	σ^2
p_{11}	1	0.4241	0.0631	0.3608	0.1015	0.0225
p_{22}		1	0.0970	0.0765	0.0071	-0.0020
μ			1	0.4967	0.8355	-0.6405
φ_1				1	0.5678	-0.3328
φ_2					1	-0.6360
σ^2						1

²³ The minimum acceptance rate while sampling h_i 's via the M-H algorithm, amounted to approximately 60%.

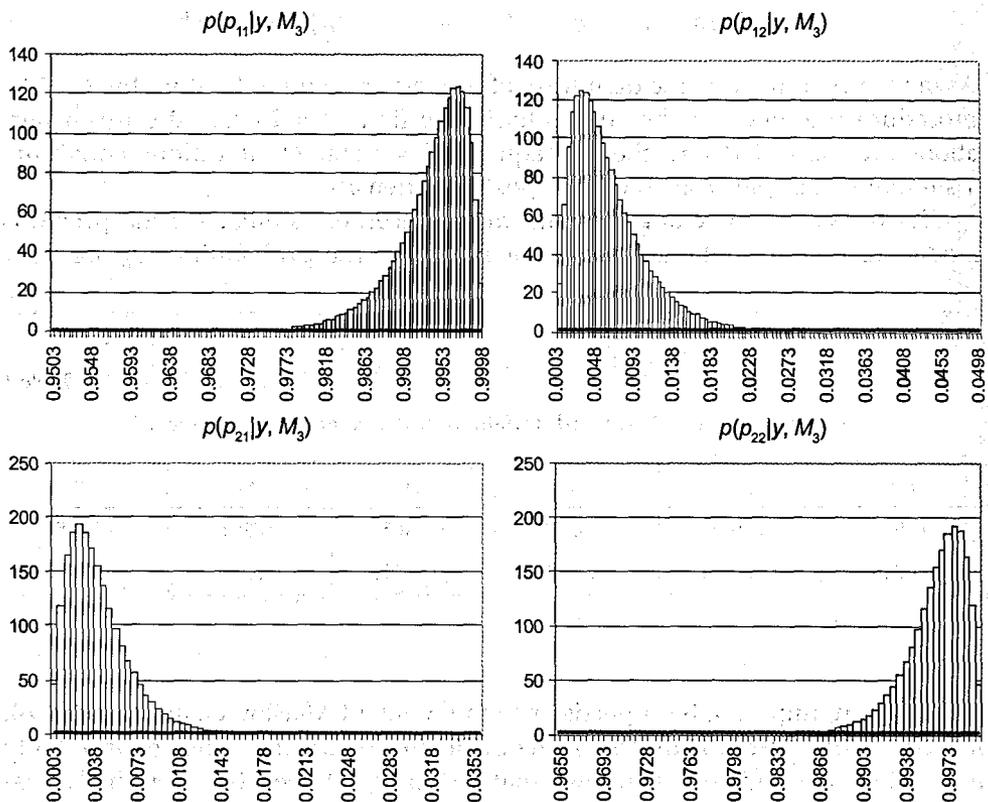


Fig. 10. Marginal prior (blue solid line) and posterior distributions of the transition probabilities in model M_3

tinct states of the economy do occur in the modelled time series. Again, the posterior correlations between the parameters appear to reject their prior independence, a reason for which is believed to be the same as in the model with a regime-changing intercept.

Marginal posterior densities of the transition probabilities resemble much those obtained for the MSSV(μ) model. Apart from a strong left asymmetry of parameter φ_1 , no other irregularities are found in the posterior distributions of the parameters (see Fig. 11).

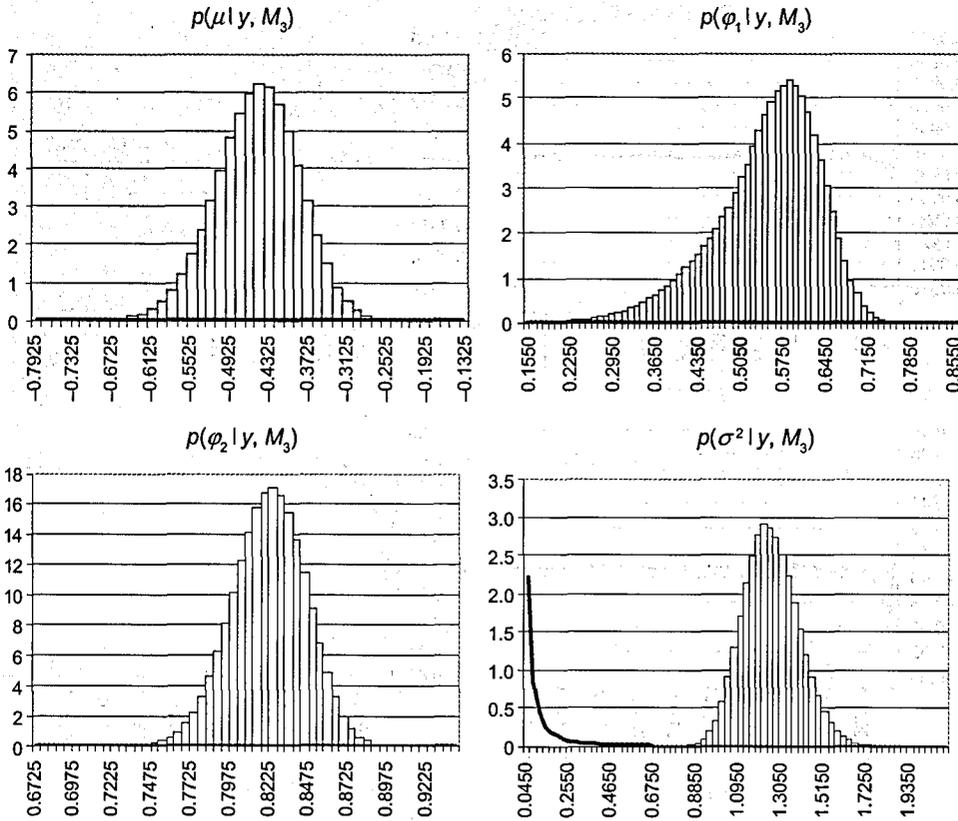


Fig. 11. Marginal prior (solid line) and posterior distributions of the log-volatility parameters in model M_3

The behaviour of the ergodic means seems to raise no concerns with regard to the convergence of the MCMC algorithm (see Fig. 12).

To analyze the validity of the prior constraints for second-order stationarity of the log-volatility process is a more demanding task than in the previous cases. Therefore, we present Figure 13, plotting the values of R_1 and R_2 , which are required to satisfy the inequalities: $R_1 < 1$ and $R_2 < 2$ (see Section 2.2). We note that only the dark area in the figure represents the set of pairs (R_1, R_2) that guarantee stationarity of the log-volatility process.²⁴ Within the region two-dimensional contours of the posterior density of random vector (R_1, R_2) are plotted. Despite the location of the latter close to the stationarity border, the data appears rather not to reject the prior stationarity restrictions.

²⁴ The stationarity region has been obtained by simulation and therefore displays slight inaccuracies.

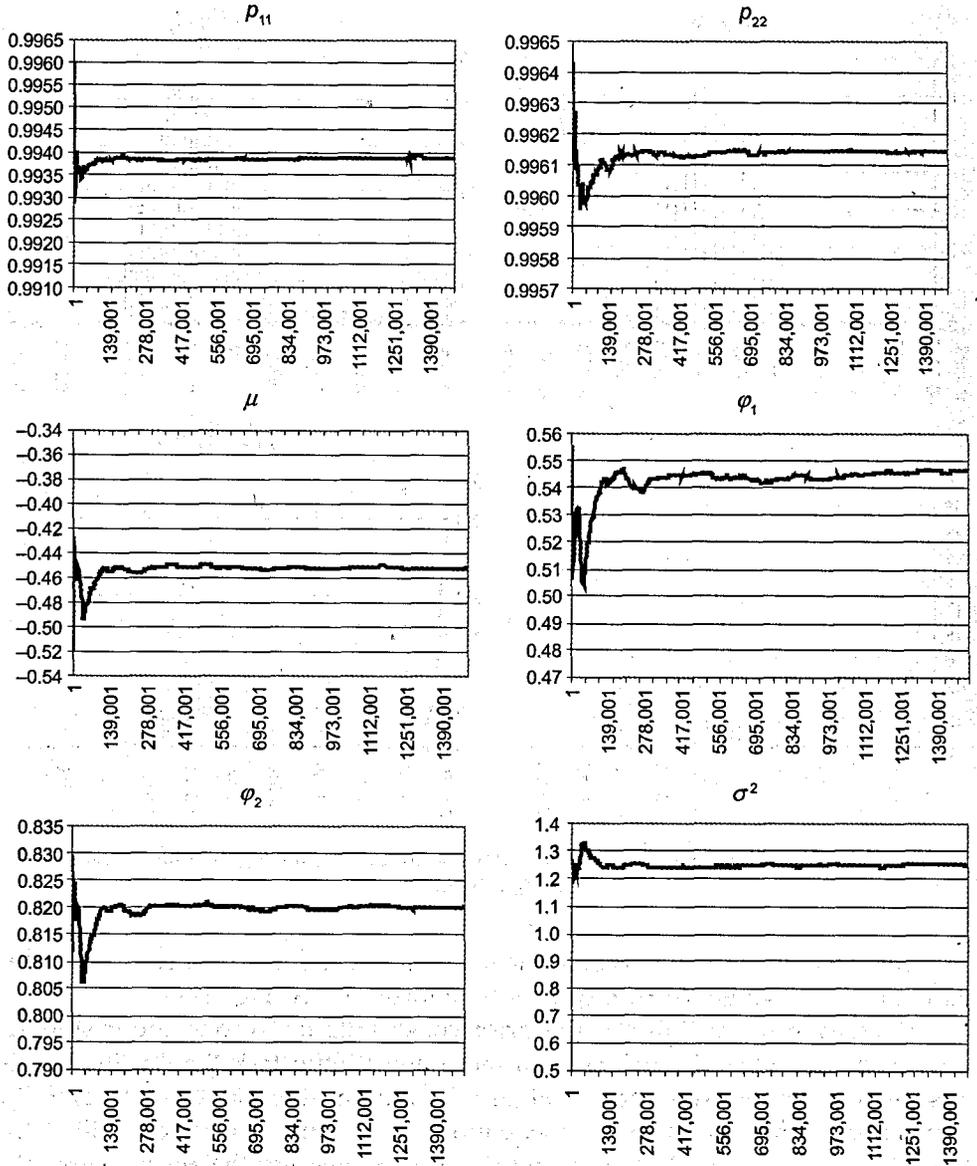


Fig. 12. Ergodic means of the parameters of model M_3 against the number of cycles

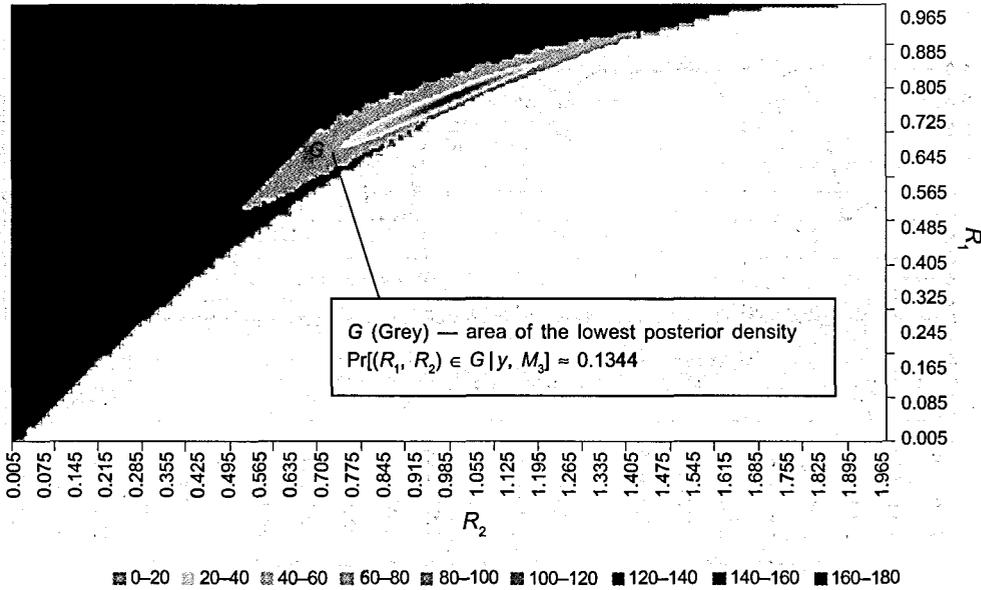


Fig. 13. Simulated region of possible values of restrictions ($R_1 < 1, R_2 < 2$) for covariance stationarity of a MSSV(ϕ) process (dark area) along with contours of their bivariate posterior distribution obtained for the data

4.5. Regime characteristics

Both regime-switching SV specifications imply existence of two distinguishable states of the economy. It is evident even more in Figure 14, depicting averaged posterior probabilities²⁵ $\Pr(S_t = 1 | y)$ in each of the two models along with the modelled time series and the averaged posterior log-volatilities, $\ln h_t$'s, extracted from model M_2 .²⁶ Unit-close values of the 1-state mean probabilities clearly correspond with the period of relatively higher volatility of the daily WIBOR1M interest rates (from about April, 2001 to September, 2004). Most of the remaining part of the sample period is definitely labelled as state '2'. There is a rather short sub-period, however, lasting from March, 2005 to September, 2005, that cannot be ascribed to any of the regimes unambiguously. It may be

²⁵ Mean posterior probabilities of state '1' have been obtained as:

$$\Pr(S_t = 1 | y) = \frac{1}{N} \sum_{q=M+1}^{N+M} I(S_t^{(q)} = 1), \quad t = 1, 2, \dots, T.$$

²⁶ The series of the averaged posterior $\ln h_t$'s only from model M_2 is presented, as it coincides quite much with the ones obtained from other specifications, i.e. M_1 and M_3 .

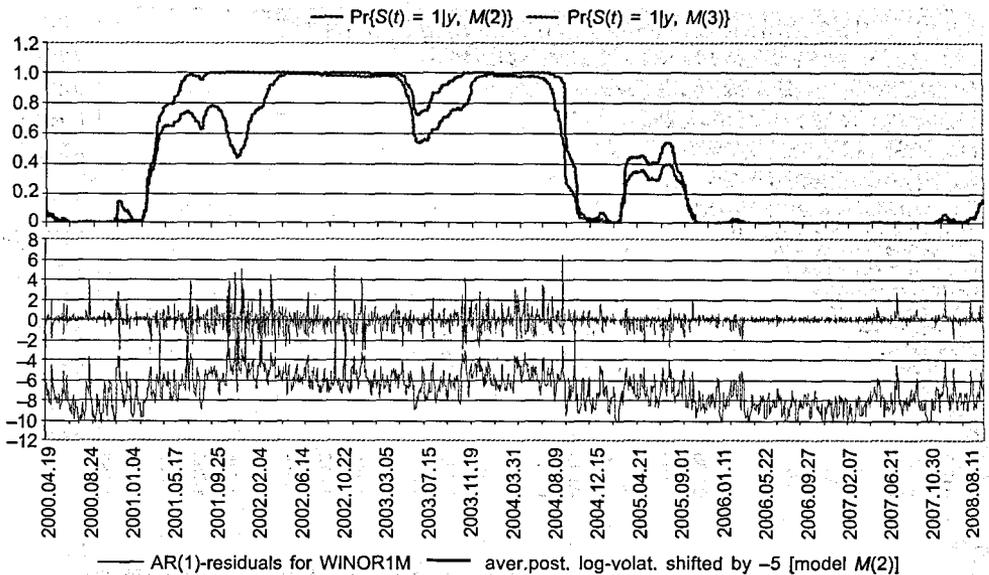


Fig. 14. Mean posterior probabilities of state '1' in models M_2 and M_3 (upper plot) along with the modelled time series (lower plot) and averaged posterior $\ln h_t$'s model M_2

case that yet another state (i.e. the one representing a medium volatility level) should be introduced to the model, yielding a three-state MSSV specification.

Posterior means of the regime characteristics (see Tab. 8) indicate that the models differentiate the two regimes in terms of either only the mean log-volatility level (model M_2) or, additionally, in terms of the state-dependent variances of the log-volatility process (model M_3), with the low-volatility state '2' featuring relatively increased 'variability of volatility'. One should note that what characterizes the expected durations of each of the states is considerable dispersion (in terms of the standard deviation) in their posterior densities featuring very long and heavy right tails (see Fig. 16). On average, the expected time of the Markov chain remaining in a particular state (once it has been achieved) differs from its posterior mean by about 805 to 5251 weekdays (see Tab. 8). As regards the ergodic probabilities, it is noticed that their posterior densities, though of a regular shape, are fairly diffused over the unit interval, therefore precluding precise inference on approximately how long²⁷ the chain remains in a particular state.²⁸ On the other hand, posterior distributions of the remaining regime characteristics (i.e. state-dependent log-volatility means and variances) evidently cluster around their posterior means, although slight asymmetries in their profiles may be observed (see Fig. 17 and 18).

²⁷ In terms of a part of the entire period over which the data is analyzed.

²⁸ We draw attention to the fact, that such an interpretation of the ergodic probabilities of a Markov chain is valid once the chain has converged to its stationary (ergodic, invariant) distribution.

Table 8

Posterior means and standard deviations (in parentheses) of selected regime characteristics in model M_2 and M_3

Model	Regime characteristics							
	π_1	π_2	Dur_1	Dur_2	E_1	E_2	V_1	V_2
M_2	0.4736	0.5264	467.60	555.61	-0.8396	-2.9871	2.4658	2.4643
[MSSV(μ)]	(0.1873)	(0.1873)	(1149.07)	(5251.38)	(0.1639)	(0.1615)	(0.1967)	(0.1964)
M_3	0.4104	0.5896	337.52	484.60	-1.0254	-2.4847	1.8301	3.8422
[MSSV(ϕ)]	(0.1799)	(0.1799)	(805.24)	(979.22)	(0.1600)	(0.1959)	(0.2428)	(0.3792)

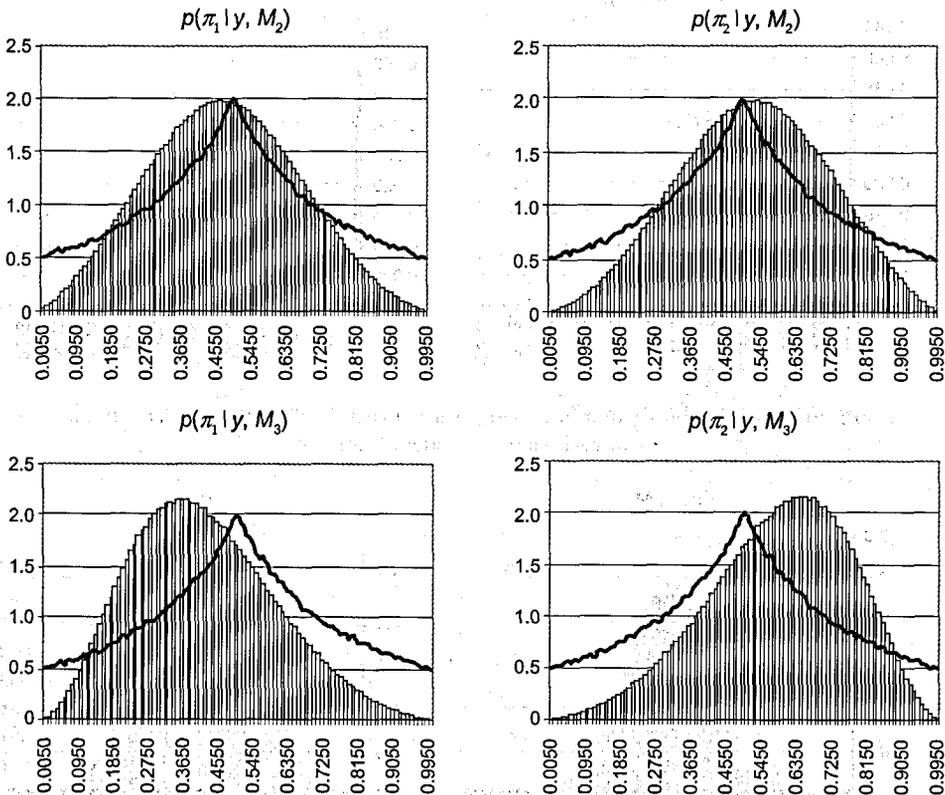


Fig. 15. Marginal prior (solid line) and posterior distributions of ergodic probabilities in model M_2 and M_3

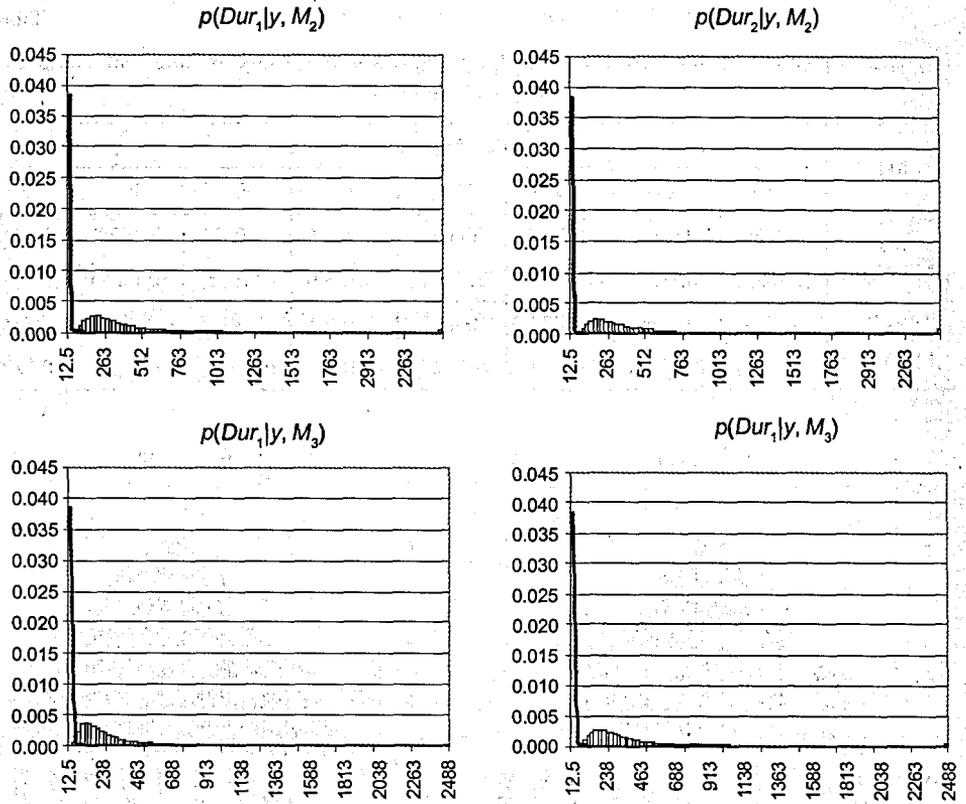


Fig. 16. Marginal prior (solid line) and posterior distributions of expected duration of each state in model M_2 and M_3

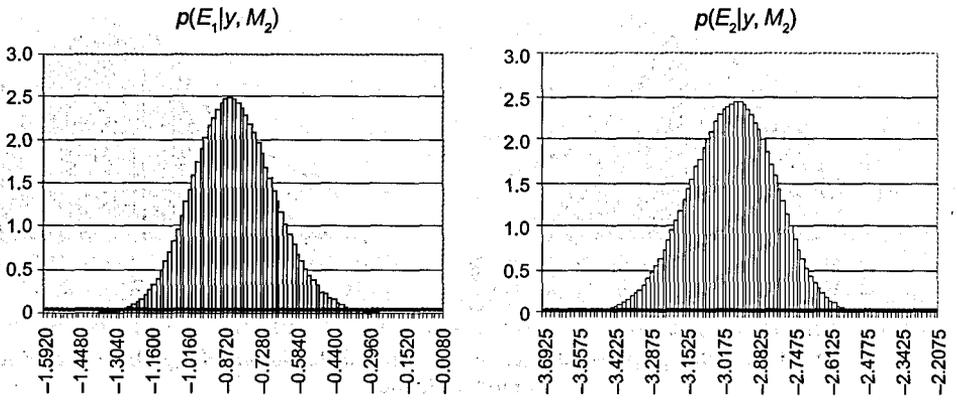


Fig. 17a. Marginal prior (solid line) and posterior distributions of state-conditional log-volatility means in model M_2

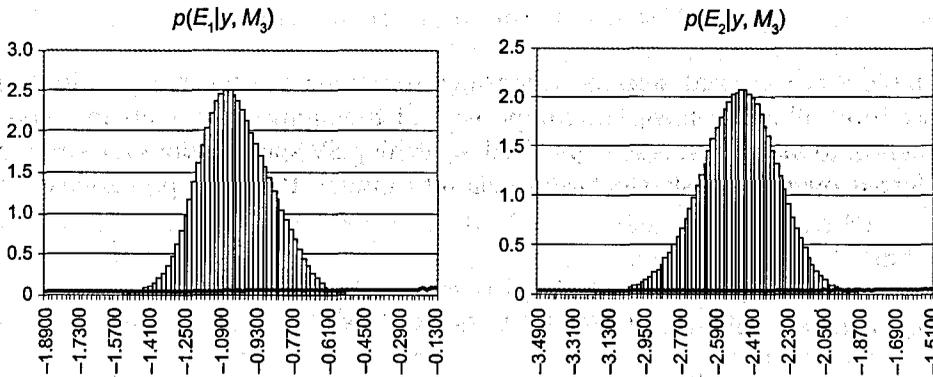


Fig. 17b. Marginal prior (solid line) and posterior distributions of state-conditional log-volatility means in model M_3

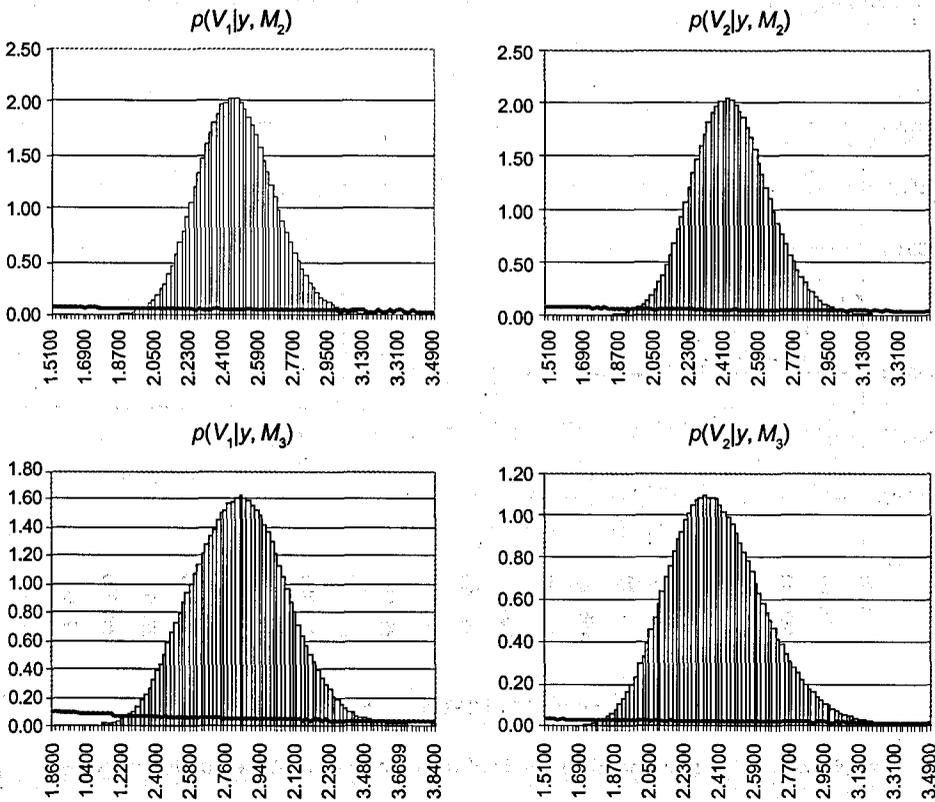


Fig. 18. Marginal prior (solid line) and posterior distributions of state-conditional log-volatility variances in model M_2 and M_3

4.6. Bayesian model comparison

In Table 9 we present selected quantities (obtained via the Newton-Raftery procedure) allowing Bayesian comparison of the analyzed models in respect of their fit to the data. It is seen that both switching SV specifications are strongly preferred over the basic stochastic volatility model. Posterior probability²⁹ of

Table 9

Logs of the marginal likelihoods along with the posterior model probabilities and logs of Bayes factors against model M_1

Model	Number of parameters	$\log_{10} \hat{p}(y M_i)$	$\Pr(M_i y)$	$\log_{10} B_{i1}$	Rank
M_1 (BSV)	3	-440.8213	6.215E-16	0	3
M_2 [MSSV(μ)]	6	-425.6148	0.999997	15.2066	1
M_3 [MSSV(φ)]	6	-431.0917	0.000003	9.7296	2

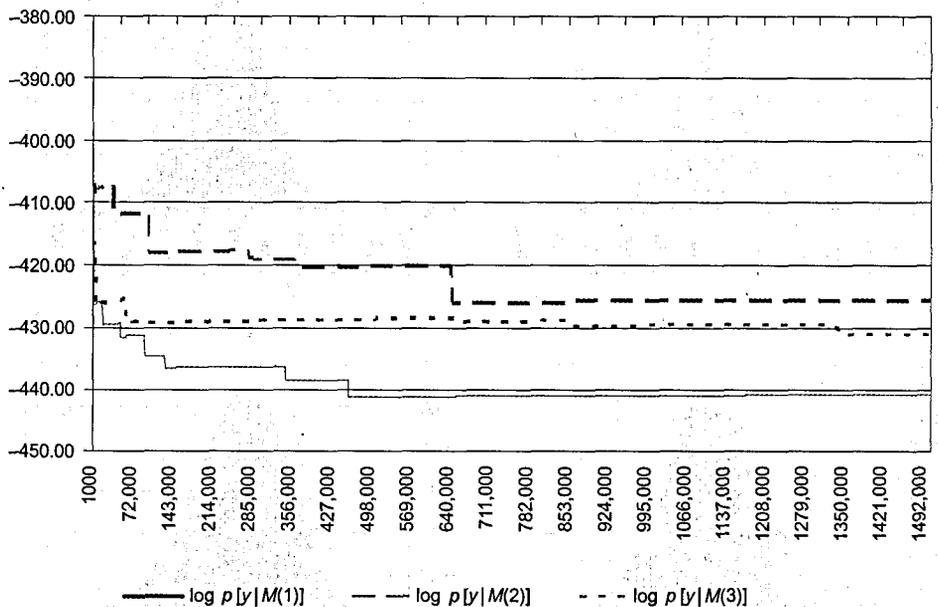


Fig. 19. Logs of the marginal likelihoods of the models against the number of cycles

²⁹ Posterior model probabilities, $\Pr(M_i | y)$, have been obtained under equal prior probabilities of the models, i.e. $\Pr(M_i) = 1/3$.

the model with a switching intercept is as much as about 10^{15} times the posterior chances of the BSV model, and about 10^5 times the chances of the other switching model. These are compelling arguments against the homogeneity (i.e. the lack of structural shifts) of the modelled time series.

Nevertheless, the results may be considered somewhat dubious in view of the notorious instability of the Newton-Raftery algorithm. Therefore, the logs of the marginal likelihood in each of the models and selected Bayes factors are plotted against the number of MCMC iterations (see Fig. 19 and 20). We observe relative stabilization of these quantities only after about 650,000 cycles. More importantly, however, the ranking of the models remains visibly unchanged throughout (see Fig. 19).

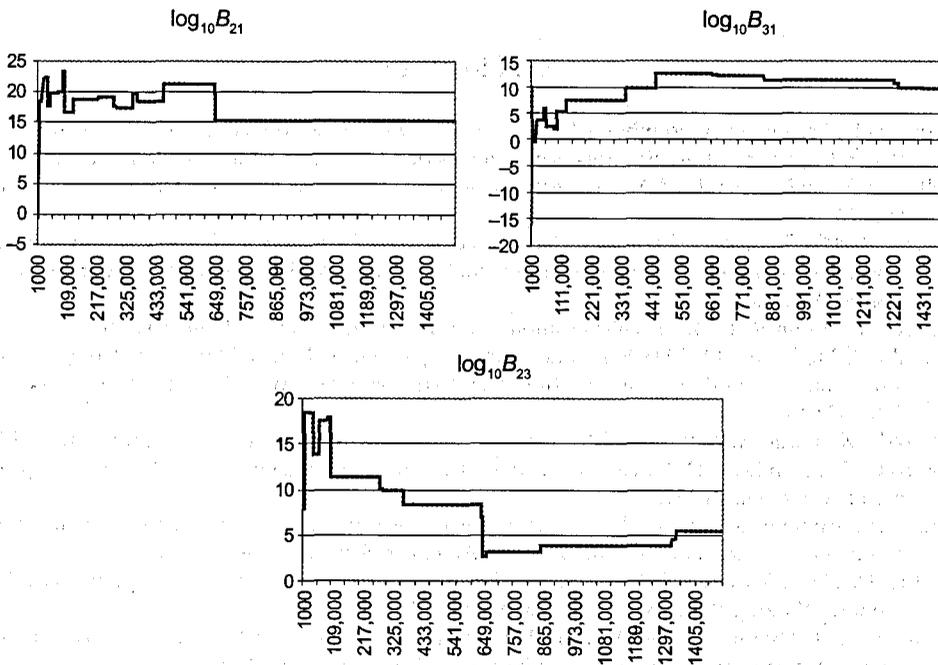


Fig. 20. Logs of Bayes factors against the number of cycles

5. CONCLUSIONS

In the paper two special cases of a general Markov switching SV model are under consideration. One of them allows discrete shifts only in the intercept, whereas the other — in the autoregression parameter of the latent log-volatility process. Both constructions are capable of accounting for sudden changes

in the mean volatility level. Hence, we aim to compare these two specifications in respect of goodness of their fit to the data.

The results of the Bayesian analysis of both switching models as well as a basic SV model provides compelling evidence against homogeneity of the series of the AR(1)-residuals for the daily WIBOR1M interest rates, as evident superiority of the switching models over the BSV construction is observed. Among them the one that features a regime-changing intercept is undoubtedly preferred the most.

The two regimes are distinguishable in terms of either only the mean log-volatility level (in the model with a switching intercept) or, additionally, the state-dependent variances of the log-volatility process (while the autoregression parameter is allowed regime-changing).

REFERENCES

- Bauwens L., Preminger A., Rombouts J. 2006. Regime Switching GARCH Models, Core Discussion Paper, Département des Sciences Économiques de l'Université catholique de Louvain.
- Bollerslev T. 1987. Generalised Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31.
- Casarin R. 2003. Bayesian Inference for Generalised Markov Switching Stochastic Volatility Models, Conference materials at the 4th International Workshop on Objective Bayesian Methodology, CNRS, Aussois.
- Carter C.K., Kohn R. 1994. On Gibbs sampling for state space models, *Biometrika* 81, 3.
- Carvalho C.M., Lopes H.F. 2006. Simulation-based sequential analysis of Markov switching stochastic volatility models, *Computational Statistics & Data Analysis*, doi: 10.1016/j.csda.2006.07.019.
- Diebold F.X., Inoue A. 2001. Long Memory and Regime Switching, *Journal of Econometrics*, 105.
- Franco C., Zakoian J.-M. 2001. Stationarity of multivariate Markov-switching ARMA models, *Journal of Econometrics* 102.
- Frühwirth-Schnatter S. 2001. Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models, *Journal of the American Statistical Association* 96.
- Gärtner D. 2007. Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, Working Paper No. 0719, University of Zurich.
- Granger C.W.J., Hyung N. 1999. Occasional Structural Breaks and Long Memory, Discussion Paper 99-14, Department of Economics, University of California, San Diego.
- Hamilton J.D. 1989. A New approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 2.
- Hwang S., Satchell S. E., Pereira P. L. V. 2003. Stochastic Volatility Models with Markov Regime Switching State Equations, *Journal of Business and Economic Statistics* 16, 2.
- Hwang S., Satchell S. E., Pereira P. L. V. 2004. How Persistent is Volatility? An Answer with Stochastic Volatility Models with Markov Regime Switching State Equations, CEA@Cass Working Paper Series, <http://www.cass.city.ac.uk/cea/index.html>
- Kalimipalli M., Susmel R. 2001. Regime-switching stochastic volatility and short-term interest rates, CEMA Working Papers, <http://ideas.repec.org/p/cem/doctra/197.html>
- Krolzig, H.-M. 1997. Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis, *Lecture Notes in Economics and Mathematical Systems*, New York/Berlin/Heidelberg: Springer.

- Newton M.A., Raftery A.E. 1994. Approximate Bayesian inference by the Weighted Likelihood Bootstrap with Discussion. *Journal of the Royal Statistical Society B* 56, 1.
- Nielsen S., Olesen J. O. 2000. Regime-switching stock returns and mean reversion, Working paper 11-2000, Institut for Nationalokonomi, <http://citeseer.ist.psu.edu>
- Shibata M., Watanabe T. 2005. Bayesian analysis of a Markov switching stochastic volatility model, *Journal of Japan Statistical Society* 35, 2.
- So M. K. P., Lam K., Li W. K. 1998. A stochastic volatility model with Markov switching, *Journal of Business and Economic Statistics* 16, 2.
- Pajor A., 2003. Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych (Stochastic Volatility Processes in Bayesian analysis of financial time series), Cracow University of Economics, Cracow 2003.
- Smith D. R., 2000. Markov-switching and stochastic volatility diffusion models for short-term interest rates, <http://citeseer.ist.psu.edu>
- Tanner M.T., Wong W. 1987. The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association* 82.

APPENDIX

Under the notation established in Section 3, the joint posterior distribution of all the unknown quantities of the MSSV model is decomposed as:

$$p(\theta, h, S | y) \propto p(y | h)p(h | S, \theta)p(S | \theta)p(\theta),$$

Individual components of the above factorization presents themselves as follows:

$$— p(y | h) = \prod_{i=1}^T f_N(y_i | 0, h_i),$$

where: $f_N(y_i | 0, h_i)$ denotes the density function of a normally distributed random variable y_i with mean and variance equal zero and h_i , respectively;

$$— p(h | S, \theta) = \prod_{i=1}^T f_{LN}(h_i | m_i, \sigma^2),$$

where: $f_{LN}(h_i | m_i, \sigma^2)$ denotes the density function of a log-normally distributed random variable h_i with the scale parameter equal σ^2 and the location parameter equal m_i given as:

$$m_i = \begin{cases} \mu_{S_i} + \varphi \ln h_{i-1} & \text{for MSSV } (\mu) \\ \mu + \varphi_{S_i} \ln h_{i-1} & \text{for MSSV } (\varphi) \end{cases};$$

$$— p(S | \theta) \propto p(S_0)p(S | \theta) = p(S_0) \prod_{i=1}^T p(S_i | S_{i-1}; \theta),$$

where $p(S_0)$ denotes the probability distribution of a discrete random variable S_0 . In the study the latter does not constitute a quantity of interest, although it is straightforward to accommodate the sampling algorithm so that inference on S_0 is available.

Under the prior structure presented in Section 3, the following conditional posterior distributions are obtained:

$$— p(p_{ii} | \theta_{-p_{ii}}, h, S, y) = f_B(p_{ii} | a_i^*, b_i^*), \text{ for } i = 1, 2,$$

where

$$\begin{aligned} a_1^* &= a_1 + n_{11}, & b_1^* &= b_1 + n_{12}, \\ a_2^* &= a_2 + n_{22}, & b_2^* &= b_2 + n_{21}, \end{aligned}$$

and

$$n_{ij} = \sum_{t=2}^T I(S_{t-1} = i)I(S_t = j);$$

$$- p(\sigma^2 | \theta_{-\sigma^2}, h, S, y) = f_{IG}(\sigma^2 | v_1^*, v_2^*),$$

where

$$v_1^* = \frac{T}{2} + v_1, \quad v_2^* = \left[\frac{1}{2} \sum_{i=1}^T (\ln h_i - m_i)^2 + \frac{1}{v_2} \right]^{-1};$$

$$- p(\beta | \theta_{-\beta}, h, S, y) = \begin{cases} f_N^{(3)}(\beta | \beta_*, \sigma^2 A_*^{-1}) I(\gamma_2 < 0) I(|\varphi| < 1) & \text{for MSSV } (\mu) \\ f_N^{(3)}(\beta | \beta_*, \sigma^2 A_*^{-1}) I(R_1 < 1) I(R_2 < 2) & \text{for MSSV } (\varphi)' \end{cases}$$

where

$$\beta_* = A_*^{-1} (\sigma^2 A_0 \beta_0 + W' \ln h), \quad A_* = \sigma^2 A_0 + W' W, \quad \ln h = (\ln h_1, \ln h_2, \dots, \ln h_T)',$$

and

$$W = \begin{bmatrix} 1 & I(S_1 = 2) & \ln h_0 \\ 1 & I(S_2 = 2) & \ln h_1 \\ \vdots & \vdots & \vdots \\ 1 & I(S_T = 2) & \ln h_{T-1} \end{bmatrix} \quad \text{for MSSV } (\mu),$$

$$W = \begin{bmatrix} 1 & I(S_1 = 1) \ln h_0 & I(S_1 = 2) \ln h_0 \\ 1 & I(S_2 = 1) \ln h_1 & I(S_2 = 2) \ln h_1 \\ \vdots & \vdots & \vdots \\ 1 & I(S_T = 1) \ln h_{T-1} & I(S_T = 2) \ln h_{T-1} \end{bmatrix} \quad \text{for MSSV } (\varphi);$$

$$p(h_i | \theta, h_{-i}, S, y) \propto \frac{1}{h_i^{1.5}} \exp\left(-\frac{y_i^2}{2h_i}\right) \exp\left(-\frac{1}{2\bar{\sigma}_i^2} (\ln h_i - w_i)^2\right),$$

where

for the MSSV(μ) model:

$$\text{for } t = 1, 2, \dots, T-1: \quad \bar{\sigma}_t^2 = \frac{\sigma^2}{1 + \varphi^2}, \quad w_t = \frac{\mu_{S_t} - \varphi \mu_{S_{t+1}} + \varphi (\ln h_{t+1} + \ln h_{t-1})}{1 + \varphi^2},$$

$$\text{for } t = T: \quad \bar{\sigma}_T^2 = \sigma^2, \quad w_T = \mu_{S_T} + \varphi \ln h_{T-1};$$

for the MSSV(φ) model:

$$\text{for } t = 1, 2, \dots, T-1: \quad \bar{\sigma}_t^2 = \frac{\sigma^2}{1 + \varphi_{S_{t+1}}^2}, \quad w_t = \frac{\mu(1 - \varphi_{S_{t+1}}) + \varphi_{S_t} \ln h_{t-1} + \varphi_{S_{t+1}} \ln h_{t+1}}{1 + \varphi_{S_{t+1}}^2},$$

$$\text{for } t = T: \quad \bar{\sigma}_T^2 = \sigma^2, \quad w_T = \mu + \varphi_{S_T} \ln h_{T-1}.$$

With regard to the conditional posterior distribution of vector S we only note here that it can be decomposed as:

$$p(S | \theta, h, y) = p(S_T | \theta, h, y) \prod_{t=1}^{T-1} p(S_t | S_{t+1}, \theta, h^t, y^t)$$

where y^t and h^t denotes the history of the observable process, and the volatility process, respectively, up to moment t . The idea has been suggested by Carter and Kohn (1994) and it is there that we refer for further details.

The MCMC sampling procedure

A full single cycle of the MCMC algorithm requires sampling each quantity of interest from its conditional posterior distribution. We employ the Gibbs procedure to sample the model parameters and vector S . For sampling the conditional volatilities, h_t 's, we adapt the Metropolis-Hastings algorithm used by Pajor (2003) in the case of non-switching SV models.

Let denote $\alpha^{(q)}$ the outcome on α from the q -th iteration³⁰ ($q = 1, 2, \dots, M, M+1, \dots, M+N$), and $\omega_{-\alpha}$ — a vector consisting of the elements of ω without its component α . Under this notation, a single full step of the sampling scheme proceeds as follows:

Step 1: sample $S^{(q+1)}$ from $p(S | \theta^{(q)}, h^{(q)}, S^{(q)}, y)$

Note: For a detailed description of the algorithm of sampling from $p(S | \theta, h, y)$ — see Carter and Kohn (1994);

Step 2: sample $p_{ii}^{(q+1)}$ from $p(p_{ii} | \theta_{-p_{ii}}^{(q)}, h^{(q)}, S^{(q+1)}, y)$, $i = 1, 2$;

Step 3: sample $\beta^{(q+1)}$ from $p(\beta | (\sigma^2)^{(q)}, p_{11}^{(q+1)}, p_{22}^{(q+1)}, h^{(q)}, S^{(q+1)}, y)$;

Step 3* — only for the MSSV(ϕ) model³¹ (*permutation sampler*; see Frühwirth-Schnatter, 2001):

→ if $\phi_1^{(q+1)} < \phi_2^{(q+1)}$ is violated, then:³²

³⁰ For $q = 0$ we obtain the set of starting values of the algorithm.

³¹ A note should be made here on sampling the model-specific parameters, β , when considering the MSSV(ϕ) model, i.e. Step 3* in the sampling scheme. In Section 2.2 the identifiability constraint $\phi_1 < \phi_2$ is imposed. To guarantee that the restriction holds an additional step, called the *permutation sampler* (see Frühwirth-Schnatter, 2001), is introduced to the Gibbs procedure. Once a new β has been sampled from its full conditional posterior distribution, we check whether the inequality $\phi_1 < \phi_2$ is violated. If so, the subscripts '1' and '2' are simply interchanged so that the restriction is valid again. Since prior to sampling β all the state variables, S_t 's, and the transition probabilities are generated, they must be 'updated' (if ϕ_1 and ϕ_2 required switching), that is all the ones and twos in vector S as well as the subscripts of the probabilities p_{ij} need to be interchanged. Relevant assumptions, theorems and proofs of the validity of such an algorithm are found in Frühwirth-Schnatter (2001).

³² By ':' we denote the substitution operator.

$\varphi_1^{(q+1)} := \varphi_2^{(q+1)}$, $\varphi_2^{(q+1)} := \varphi_1^{(q+1)}$ (so that $\varphi_1^{(q+1)} < \varphi_2^{(q+1)}$ is guaranteed),

$S^{(q+1)} := 3l_T - S^{(q)}$ where $l_T' = (1, 1, \dots, 1)_{(1 \times T)}$,

$p_{ij}^{(q+1)} := p_{3-i, 3-j}^{(q)}$ for $i, j = 1, 2$;

Step 4: sample $(\sigma^2)^{(q+1)}$ from $p(\sigma^2 | \theta_{-\sigma^2}^{(q+1)}, h^{(q)}, S^{(q+1)}, y)$;

Step 5: (the Metropolis-Hastings step): sample each $h_t^{(q+1)}$ from

$p(h_t | \theta^{(q+1)}, h_{(1:t-1)}^{(q+1)}, h_{(t+1:T)}^{(q)}, S^{(q+1)}, y)$, where $h_{(s:t)} = (h_s, h_{s+1}, \dots, h_t)$

with $s \leq t$.