Central European Journal of Economic Modelling and Econometrics

Markov Switching In-Mean Effect. Bayesian Analysis in Stochastic Volatility Framework

Łukasz Kwiatkowski*

Submitted: 31.03.2010, Accepted: 7.07.2010

Abstract

In the study we introduce an extension to a stochastic volatility in mean model (SV-M), allowing for discrete regime switches in the risk premium parameter. The logic behind the idea is that neglecting a possibly regime-changing nature of the relation between the current volatility (conditional standard deviation) and asset return within an ordinary SV-M specification may lead to spurious insignificance of the risk premium parameter (as being 'averaged out' over the regimes). Therefore, we allow the volatility-in-mean effect to switch over different regimes according to a discrete homogeneous two-state Markov chain. We treat the new specification within the Bayesian framework, which allows to fully account for the uncertainty of model parameters, latent conditional variances and hidden Markov chain state variables. Standard Markov Chain Monte Carlo methods, including the Gibbs sampler and the Metropolis-Hastings algorithm, are adapted to estimate the model and to obtain predictive densities of selected quantities. Presented methodology is applied to analyse series of the Warsaw Stock Exchange index (WIG) and its sectoral subindices. Although rare, once spotted the switching in-mean effect substantially enhances the model fit to the data, as measured by the value of the marginal data density.

Keywords: Markov switching, stochastic volatility, risk premium, in-mean effect, Bayesian analysis

JEL Classification: C50, C11, C22.

*Cracow University of Economics, e-mail: kwiatkol@uek.krakow.pl

1 Introduction

Although the idea of a compensation to a risk-averse investor for holding a risky asset appears theoretically sound and intuitively appealing, extensive empirical research has been unable to establish a convincing positive relationship between the expected returns and volatility. Particularly, the abounding literature exploiting various specifications of the ARCH-in-Mean (ARCH-M) structure of Engle, Lilien, Robins (1987), has provided evidence of both a positive and a negative risk-return tradeoff; see, for instance, Nelson (1991), Glosten, Jagannathan, Runkle (1993), Scruggs (1998) and N'dri (2008) for more references. Such perplexing results have spurred the researchers to investigate a GARCH-in-Mean (GARCH-M) model with the risk premium parameter following a random walk in order to detect a possibly time-varying pattern of the risk-return relationship; see Chou, Engle, Kane (1992), Fiszeder and Kwiatkowski (2005a), (2005b). In the meantime, Backus and Gregory (1993) have made an attempt to unravel the issue of the sign of the risk premium parameter within some theoretical economic model and concluded that the theory does not guarantee the relation between conditional mean of excess returns (the returns net of the risk-free rate) and time-varying volatility to be either an increasing or even monotonic function of the latter. In search of a more statistical premise of the risk premium Osiewalski and Pipień (2000) and Pipień and Osiewalski (2001) develop yet another GARCH-M specification, in which the noise term in the observation equation is modelled with a skewed t-distribution. Their structure is able to separate potentially two formal, statistical sources of the in-mean effect: skewness of the conditional distribution and the location of its mode. The empirical results, however, yet again provided evidence of insignificant GARCH-in-Mean effect. Quite apart from the GARCH literature, there has appeared a sole study in which a stochastic volatility (SV) counterpart of the GARCH-M structure has been proposed, i.e. the SV-M model of Koopman and Hol Uspensky (2002). Most probable reason for such a sizeable disproportion in the attention paid to the two different model classes is that the estimation of the SV structure is less straightforward, on account of the presence of a hidden volatility process.

In the paper we revisit the volatility-in-mean effect, yet in search of its possibly switching nature. We formulate a new stochastic volatility model, incorporating the interdependence between the current volatility and stock market return, and allowing for its discrete shifts over time. There is a simple line of reasoning behind our concept. Namely, if the 'true' risk premium features different states (regimes), then considering an in-mean model with a constant risk premium parameter may lead to its spurious insignificance. Negligence of the regime-changing pattern may lead to the results that are somewhat 'averaged out' over the regimes. Therefore, we suggest that the risk-return parameter be subject to discrete switches, governed by a two-state, homogenous and ergodic Markov process, much in the vein of the seminal work of Hamilton (1989). The final product constitutes a Stochastic Volatility Markov Switching in Mean model, or SV-MS-M in short.

The main thrust of our work is to find out whether the risk-return interrelationship may display switching pattern. Therefore, three SV specifications are under study: a basic stochastic volatility model, a SV-M model similar to the one introduced by Koopman and Hol Uspensky (2002), and, finally, a new regime-switching construction. The latter, if the switching phenomenon is revealed, further serves to characterize the discerned regimes. Lastly, the three structures are examined in respect of their prediction abilities.

As regards the estimation of each model, we resort to the Bayesian inference rather than the classical (i.e. non-Bayesian) approach. Although a maximum likelihood treatment of the SV-MS-M model seems feasible (extending, for example, the technique presented by Koopman and Hol Uspensky 2002, so as to account for the presence of an underlying hidden Markov chain), the Bayesian methodology appears more attractive on several counts. First and foremost, it does take the uncertainty inherent to all the unknown quantities of the model into account. In the regime-switching models (or, generally, the mixture models) the argument gains even more weight (as aptly recognized and explicated by Gärtner 2007) due to uncertainty inherent to the mixture components indices. Secondly, Bayesian treatment of the latent variables (the conditional volatilities and state variables) as additional parameters is far more tractable than the way the classical approach copes with them. Moreover, an appropriate choice of the prior distributions (so that they are conjugate to their posterior conditional counterparts), enables us to employ standard MCMC techniques, such as the Gibbs sampler and the Metropolis-Hastings algorithm, to sample from the joint posterior and predictive distributions. Finally, the Bayesian perspective provides a natural approach to model comparison in terms of the in-sample fit, by means of the value of the marginal data density, evaluated for each considered specification.

There have been several previous attempts to introduce hidden Markov process into stochastic volatility structure. These, however, have always aimed at modelling discrete changes solely in the log-volatility parameters, mostly the intercept, see So, Lam, Li (1998), Smith (2000), Kalimipalli and Susmel (2001), Valls Pereira (2004), Casarin (2003), Shibata and Watanabe (2005), Carvalho and Lopes (2006), Kwiatkowski (2009a), (2009b). On the other hand, the present paper employs the idea of the Markov switching mechanism in a different context, hence contributing to the vein of combining the Markov process with SV structures.

The remainder of the paper is organized as follows. In the following section we present the SV-MS-M model in more detail. The estimation procedure, Bayesian forecasting and model comparison are discussed in Section 3. Presented methodology is illustrated with an empirical study in Section 4, in which ten time series of daily logarithmic rates of return are under consideration: returns on the Warsaw Stock Exchange index (WIG) and its sectoral subindices. Finally, Section 5 concludes.

2 Stochastic Volatility Markov Switching In Mean model

Let the sequence $\{S_t, t \in \mathbb{Z}\}$, with $\mathbb{Z}$ denoting the set of integers, constitute a two-state ergodic Markov chain. Each discrete-valued random variable S_t takes on value of 1 or 2, i.e. $S_t \in Q = \{1, 2\}$, and signifies the index of the 'current' regime (i.e. the one at time t). The transition probabilities are defined as $p_{ij} = \Pr(S_t = j | S_{t-1} = i)$ with $i, j \in Q$. Although it would be possible to introduce dependence of the transition probabilities on some variables (the lagged modelled rates of return, for example), we assume the chain to be homogenous.
Combining the switching mechanism presented above with a typical stochastic volatility in-mean structure, we obtain the following definition of a two-state SV-MS-M model.

Definition 1 *A stochastic process $\{y_t, t \in \mathbb{Z}\}$ follows a two-state SV-MS-M process if for each $t \in \mathbb{Z}$ the following conditions hold:*

$$y_t = \delta_0 + \delta_1 y_{t-1} + \gamma_{S_t} g(h_t) + \varepsilon_t \sqrt{h_t}, \tag{1}$$

$$\ln h_t = \mu + \varphi \ln h_{t-1} + \sigma \eta_t, \tag{2}$$

$$\left\{ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}, t \in \mathbb{Z} \right\} \sim iiN^{(2)} \left(\mathbf{0}_{(2\times 1)}, I_2 \right), \tag{3}$$

with $g(h_t)$ being some increasing function of the conditional variance, h_t, and S_t representing a two-state, homogenous and ergodic Markov process defined as above.

Equation (1), hereafter being referred to as the observation equation, defines a simple first-order autoregression on y_t, completed with an additional explanatory variable, $g(h_t)$, and the innovation term $\varepsilon_t \sqrt{h_t}$. Taking (2) and (3) into consideration, the innovations follow a basic stochastic volatility process, the very first foundations of which have been laid by Clark (1973). As already mentioned, h_t is a conditional variance of y_t, once conditioning is made with respect to a σ-field generated by the lagged y_t's, the current noise term η_t and the current state variable S_t, i.e. $Var(y_t | F_{t-1}, \eta_t, S_t) = h_t$, where F_{t-1} is the past information about the process $\{y_t, t \in \mathbb{Z}\}$ up to time $t-1$ (formally being the smallest σ-field generated by η_r and $\varepsilon_r, r \leq t-1$).
Specification formulated in Definition 1 is meant to model not only the dynamics of the financial asset rates of return along with its time-varying volatility, but also the risk-return relationship. Regressing contemporaneous returns on some increasing function of conditional variance h_t stays much in the spirit of Engle et al. (1987), who introduced the ARCH-in-Mean (ARCH-M) model to capture time-varying term structure premium. Note that the SV-MS-M structure encompasses the SV-in-Mean model (once $\gamma_1 = \gamma_2 = \gamma$, although p_{ij}'s are then unidentified) and the SV model

$(\gamma_1 = \gamma_2 = \gamma = 0)$, both of which are of interest in the empirical study.
As regards the choice of the function $g(h_t)$, three different forms are commonly exploited in the literature: the logarithm or square root of h_t, or the conditional variance itself; see Engle, Lilien, Robins (1987). In our work we restrict ourselves to the second option, i.e. $g(h_t) = \sqrt{h_t}$. The main reason behind such a choice is existence of the following equivalent formulation of the observation equation:

$$\begin{array}{rcl} y_t & = & \delta_0 + \delta_1 y_{t-1} + \gamma_{S_t}\sqrt{h_t} + \varepsilon_t\sqrt{h_t} \\ & \Leftrightarrow & y_t = \delta_0 + \delta_1 y_{t-1} + \left(\gamma_{S_t} + \varepsilon_t\right)\sqrt{h_t} \\ & \Leftrightarrow & y_t = \delta_0 + \delta_1 y_{t-1} + \xi_t\sqrt{h_t} \end{array} \tag{4}$$

with $\xi_t = \gamma_{S_t} + \varepsilon_t$. Since the disturbances $\{\varepsilon_t, t \in \mathbb{Z}\}$ are independent and Normally distributed with zero mean and unit variance, it follows that

$$\xi_t | S_t \sim iiN\left(\gamma_{S_t}, 1\right) \tag{5}$$

and

$$p\left(\xi_t|\theta\right) = \pi_1 f_N\left(\xi_t|\gamma_1, 1\right) + \pi_2 f_N\left(\xi_t|\gamma_2, 1\right) \tag{6}$$

with $f_N\left(\cdot|a, b\right)$ denoting density of the univariate Normal distribution with mean a and variance b, and π_i being the ergodic probabilities of the chain $\{S_t, t \in \mathbb{Z}\}$, defined as

$$\pi_i = \Pr\left(S_t = i\right) = \frac{1 - p_{3-i,3-i}}{2 - p_{11} - p_{22}}, \tag{7}$$

for $i = 1, 2$.
According to (4)-(6), the SV-MS-M process can be represented as a simple stochastic volatility process with no explicit relationship between contemporaneous return and conditional variance, yet with the error term ξ_t constituting a two-component Markov mixture of conditionally Normal distributions with state-dependent mean γ_{S_t} and unit variance.
We find it informative to study the moment structure of the random variable ξ_t. It is straightforward to show the following:

$$E\left(\xi_t\right) = \pi_1\gamma_1 + \pi_2\gamma_2, \tag{8}$$

$$Var\left(\xi_t\right) = 1 + \pi_1\gamma_1^2 + \pi_2\gamma_2^2 - \left(\pi_1\gamma_1 + \pi_2\gamma_2\right)^2, \tag{9}$$

$$\begin{array}{c} E\left(\xi_t - E\left(\xi_t\right)\right)^3 = \gamma_1^3\left(\pi_1 - 3\pi_1^2 + 2\pi_1^3\right) + \gamma_2^3\left(\pi_2 - 3\pi_2^2 + 2\pi_2^3\right) \\ +3\pi_1\pi_2\gamma_1\gamma_2\left(\pi_1 - \pi_2\right)\left(\gamma_1 - \gamma_2\right). \end{array} \tag{10}$$

Apart from a possibly non-zero mean (see formula (8)), the unconditional distribution of the error term ξ_t may also display asymmetry, since the skewness coefficient based on the third central moment presented in (10), i.e.

$$Sk\left(\xi_t\right) = \frac{E\left(\xi_t - E\left(\xi_t\right)\right)^3}{\left(Var\left(\xi_t\right)\right)^{1.5}},$$

is generally non-zero. On the other hand, as conditionally upon the current state ξ_t is Normally distributed (see (5)), the distribution of ξ_t given S_t reveals no asymmetry at all. Finally, we note that

$$\begin{aligned} p\left(\xi_t|S_{t-1}=k,\theta\right) &= E_{S_t}\left(\xi_t|S_t,S_{t-1}=k,\theta\right) \\ &= \sum_{i=1}^{2}\Pr\left(S_t=i|S_{t-1}=k,\theta\right)\cdot p\left(\xi_t|S_t=i,S_{t-1}=k,\theta\right) \\ &= p_{k1}\cdot p\left(\xi_t|S_t=1,\theta\right)+p_{k2}\cdot p\left(\xi_t|S_t=2,\theta\right) \\ &= p_{k1}f_N\left(\xi_t|\gamma_1,1\right)+p_{k2}f_N\left(\xi_t|\gamma_2,1\right), \end{aligned}$$

which is similar to (6) with the only difference residing in the 'weights' of the mixture (i.e. probabilities). It follows that conditionally upon the lagged regime, S_{t-1}, the distribution of ξ_t reveals a similar moment structure as the unconditional distribution (see formulae (8)-(10)). Namely, we have:

$$E\left(\xi_t|S_{t-1}=i\right)=p_{i1}\gamma_1+p_{i2}\gamma_2, \tag{11}$$

$$Var\left(\xi_t|S_{t-1}=i\right)=1+p_{i1}\gamma_1^2+p_{i2}\gamma_2^2-\left(p_{i1}\gamma_1+p_{i2}\gamma_2\right)^2, \tag{12}$$

$$\begin{aligned} E\left[\left(\xi_t-E\left(\xi_t|S_{t-1}=i\right)\right)^3|S_{t-1}=i\right] &= \gamma_1^3\left(p_{i1}-3p_{i1}^2+2p_{i1}^3\right) \\ +\gamma_2^3\left(p_{i2}-3p_{i2}^2+2p_{i2}^3\right)+3p_{i1}p_{i2}\gamma_1\gamma_2\left(p_{i1}-p_{i2}\right)\left(\gamma_1-\gamma_2\right). \end{aligned} \tag{13}$$

Analogously, the distribution of ξ_t given S_{t-1} may display asymmetry, for no longer the skewness coefficient:

$$Sk\left(\xi_t|S_{t-1}\right)\equiv\frac{E\left(\left(\xi_t-E\left(\xi_t|S_{t-1}\right)\right)^3|S_{t-1}\right)}{\left(Var\left(\xi_t|S_{t-1}\right)\right)^{1.5}}$$

has to be zero.

In view of the above it is legitimate to conclude that a regime-switching in-mean effect in the SV-MS-M model is tantamount to discrete shifts in the mean of the error term ξ_t, with the latter being a component of the stochastic volatility specification shown in (4). Moreover, as 'flexible' as a two-component Markov mixture may be in respect of its (third-order) moment structure, it may be the case that apparent switches in the risk premium parameter, γ_{S_t}, could be indicative of an asymmetric distribution of ξ_t and, thereby, a skewed conditional distribution of the observed variable, y_t. Therefore, what seems to merit further research is to compare two models in terms of their in-sample fit: the SV-MS-M specification introduced in the paper, and a simple SV structure with the observation equation presented in (4), yet with the error term ξ_t following a skewed Normal or Student-t distribution, similarly as presented by Osiewalski and Pipień (2000), and Pipień and Osiewalski (2001) in the GARCH-M setting. However, the issue goes beyond the scope of the current study and will not be addressed here. Bearing in mind the potential skewness of the disturbance term ξ_t (both unconditionally and conditionally upon the lagged regime), in the empirical

part we also report on the posterior results for the relevant skewness coefficients, i.e. $Sk(\xi_t)$ and $Sk(\xi_t|S_{t-1}=i)$. Additionally, some quantities typically considered for Markov chains will be of interest, including the ergodic probabilities defined in (7), and the expected durations of each state, calculated as $Dur_i = \frac{1}{1-p_{ii}}$; see Hamilton (1989), p. 374.

3 Bayesian estimation and forecasting for the SV-MS-M model

3.1 General remarks

In this section we provide a concise description of the estimation and forecasting methodology applied in our study. Only the most general model, i.e. the SV-MS-M structure, is considered here, since it also encompasses the two other specifications under study, i.e. the SV and SV-M model. (For a detailed presentation of a Bayesian treatment of the basic SV model we refer to Jacquier, Polson, Rossi (1994) and Pajor (2003), although there are some differences between our prior structure and the one assumed in the cited papers.) The only differences between the three models reside in the observation equation (see Table 1), or, equivalently, in the distribution of ξ_t. We note that our specification of the SV-M model differs slightly from the one studied by Koopman and Hol Uspensky (2002), who regress the returns on the conditional variances rather than the standard deviations.

Table 1: Mean and volatility equation specifications for the analysed models

Model	Observation equation	Log-volatility equation
M_1: SV	$y_t = \delta_0 + \delta_1 y_{t-1} + \varepsilon_t\sqrt{h_t}$	$\ln h_t = \mu + \varphi \ln_{h_{t-1}} + \sigma\eta_t$
M_2: SV-M	$y_t = \delta_0 + \delta_1 y_{t-1} + \gamma\sqrt{h_t} + \varepsilon_t\sqrt{h_t}$	
M_3: SV-MS-M	$y_t = \delta_0 + \delta_1 y_{t-1} + \gamma_{S_t}\sqrt{h_t} + \varepsilon_t\sqrt{h_t}$	

We establish the notation first. Let $y = (y_1, y_2, \ldots, y_T)' \in Y \subseteq \mathbb{R}^T$ denote the vector of T observations on the modelled logarithmic rates of return, $h = (h_1, h_2, \ldots, h_T)' \in H \subseteq \mathbb{R}_+^T$ - the vector of the latent conditional variances, and $S = (S_1, S_2, \ldots, S_T)' \in Q^T$ - the vector of the hidden Markov chain state variables. The Markov switching risk premium parameter, γ_{S_t}, is parameterized as

$$\gamma_{S_t} = \gamma_1 + \tau I(S_t = 2),$$

with $\gamma_1 \in \mathbb{R}$, $\tau \leq 0$ and $I(\cdot)$ denoting the indicator function taking on the value of one if the condition in the parentheses is satisfied, and zero otherwise. Consequently, the in-mean effect in the second regime is lower than the one in the first state. Although it is γ_1 and γ_2 that are of interest, we consider the estimation in terms of γ_1 and τ,

noting that the posterior density of γ_2 is induced by the two latter. Parameters of the SV-MS-M model are arrayed in $\theta = \left(\delta', \gamma_1, \tau, p_{11}, p_{22}, \beta', \sigma^2\right)' \in \Theta \subset \mathbb{R}^9$, where $\delta = (\delta_0, \delta_1)'$ and $\beta = (\mu, \varphi)'$. Note that δ, β and σ^2 are common to all three models under consideration (see Table 1).
Inference on all the unknown quantities of the model is based on the joint posterior distribution of $\omega = (\theta', h', S')' \in \Omega = \Theta \times H \times Q^T$, represented by its density

$$p\left(\theta, h, S|y\right) \propto p\left(y|\theta, h, S\right) p\left(S|\theta\right) p\left(h|\theta\right) p\left(\theta\right) \tag{14}$$

where

$$p\left(y|\theta, h, S\right) \propto p(y_0) p\left(y|\theta, h, S\right) = p(y_0) \prod_{t=1}^{T} f_N\left(y_t|\delta_0 + \delta_1 y_{t-1} + \gamma_{S_t}\sqrt{h_t}, h_t\right), \tag{15}$$

$$p\left(S|\theta\right) \propto p(S_0) p\left(S|\theta\right) = p(S_0) \prod_{t=1}^{T} p\left(S_t|S_{t-1}, \theta\right) = p(S_0) \prod_{t=1}^{T} p_{S_{t-1}, S_t}, \tag{16}$$

$$p\left(h|\theta\right) \propto p(h_0) p\left(h|\theta\right) = p(h_0) \prod_{t=1}^{T} \left(\frac{1}{h_t} f_N\left(\ln h_t|\mu + \varphi \ln h_{t-1}, \sigma^2\right)\right), \tag{17}$$

$$p\left(\theta\right) = p(\delta) p(\gamma_1) p(\tau) p(p_{11}) p(p_{22}) p(\beta) p(\sigma^2). \tag{18}$$

In our work we assume that the initial values in (15) and (17), i.e. y_0 and h_0, are fixed and equal to the first presample value of the modelled series and 1, respectively. Therefore, the analysis is conducted conditionally upon these initial conditions, with it being suppressed in the notation henceforth. For the state of the switching mechanism at time $t = 0$, i.e. S_0, one could specify some arbitrary discrete-valued distribution, characterized by either its own parameters subject to statistical inference themselves, or the ergodic probabilities of the underlying Markov chain (should S_0 be Bernoulli-distributed with the success probability equal to either of the ergodic probabilities presented in formula (7)). However, the first solution would require introduction of additional parameters, whereas the second option would lead to less tractable conditional posterior densities, utilized in the sampling algorithm. Therefore, for simplicity, we assume that $\Pr(S_0{=}1) = \Pr(S_0{=}2) = 0.5$. We have some belief that the posterior results are robust to the choice of a particular way S_0 is treated.

3.2 Prior structure

According to (18) mutual prior independence of the parameters is assumed. It also applies to the individual elements of vector $\delta = (\delta_0, \delta_1)'$ and $\beta = (\mu, \varphi)'$, for both of which the following truncated bivariate Normal distributions with zero correlations are specified:

$$p\left(\delta\right) \propto f_N^{(2)}\left(\delta|d_0, C_0^{-1}\right) I\left(|\delta_1| < 1\right), \; d_0 = \mathbf{0}_{(2\times 1)}, \; C_0 = 0.01 \cdot I_2, \tag{19}$$

$$p(\beta) \propto f_N^{(2)}\left(\beta|b_0, A_0^{-1}\right) I(|\varphi| < 1),\ b_0 = \mathbf{0}_{(2\times 1)},\ A_0 = 0.01 \cdot I_2, \tag{20}$$

where I_n stands for a $(n \times n)$-sized identity matrix and $f_N^{(m)}$ - the density of the m-variate Normal distribution. Truncations of the parameter space, made by the inequality restrictions in (19) and (20), are meant to guarantee non-explosiveness of the SV-MS-M process (or its second-order stationarity once the limiting case of $t \to \infty$ is considered). However, we stress that beyond a basic intuition no formal argument underpins these restrictions. That is, we cannot be sure if the non-explosiveness restrictions that hold separately for the first-order autoregression in the observation equation and the one defining the log-volatility process, once imposed jointly for the SV-MS-M structure, constitute either a sufficient or necessary regularity condition. The issue merits further investigation.

To complete the prior structure we specify distributions for σ^2 and the model-specific parameters, i.e. p_{11}, p_{22}, γ_1 and τ, as follows:

$$p\left(\sigma^2\right) = f_{IG}\left(\sigma^2|\nu_1, \nu_2\right),\ \nu_1 = 1.5, \quad \nu_2 = 4, \tag{21}$$

$$p(p_{ii}) = f_B(p_{ii}|a_i, b_i),\ a_i = b_i = 1,\ i = 1, 2, \tag{22}$$

$$p(\gamma_1, \tau) \propto f_N^{(2)}\left(\gamma_1, \tau|\lambda_0, \Lambda_0^{-1}\right) I(\tau < 0),\ \lambda_0 = \mathbf{0}_{(2\times 1)},\ \Lambda_0 = I_2, \tag{23}$$

with f_{IG} and f_B denoting densities of the Inverse-Gamma and Beta distribution, respectively. The density in (21) is parameterized so that $E(\sigma^2) = 0.5$ ($Var(\sigma^2)$ does not exist), and the precision σ^{-2} is a Gamma-distributed random variable with $E(\sigma^{-2}) = 6$ and $Var(\sigma^{-2}) = 24$. As informative as these may appear, the prior information on σ^2 (and, equally, on σ^{-2}) is largely dominated by the data (comparing the moments of the prior and the posterior distribution; see Table 7). For the transition probabilities p_{ii} the Uniform distribution is specified, as a special case of the Beta family. We note that the truncated bivariate standard Normal in (23) induces a prior for $\gamma_2 = \gamma_1 + \tau$ with the mean:

$$E(\gamma_2) = E(\tau) = -2\phi(0) \approx -0.798,$$

and the variance:

$$Var(\gamma_2) = 2\left[1 - 2\left(\phi(0)\right)^2\right] \approx 1.363,$$

where $\phi(x) \equiv f_N(x|0, 1)$.

Densities (19)-(21) are shared across all three model specifications, i.e. the SV, SV-M and SV-MS-M model. In the case of the risk premium parameter within the SV-M structure, i.e. γ, a standard Normal prior is assumed. The prior structure exposed in (19)-(23) is intended to represent our vague beliefs as of the model parameters. However little diffuse the standard Normal distributions for γ_1 and τ may appear (see expression (23)), they reflect our prior conviction as of the magnitude of the volatility-in-mean effect.

3.3 Sampling algorithm

To estimate the considered models we resort to common MCMC techniques, including the Gibbs sampler and the Metropolis-Hastings algorithm, although an additional procedure is required to handle the latent Markov state variables, S_t's, $t = 1, 2, \ldots, T$. Generating a pseudo-random sample from the joint posterior (14) can be divided into three major steps, in which the three: the parameters, θ, conditional variances, h_t's, and state variables, S_t's, are sampled from their full conditional posterior distributions. Specifying conjugate priors, as the ones in (21)-(23), is very conducive to sampling each component of θ from its conditional posterior, since the latter belongs to the same distribution family as its prior counterpart. Below, the full conditional posteriors for the components of θ are presented.

1. $p\left(\delta, \gamma_1, \tau | \theta_{-(\delta,\gamma_1,\tau)}, h, S, y\right) \propto f_N^{(4)}\left(\delta, \gamma_1, \tau | g_*, G_*^{-1}\right) I(|\delta_1| < 1) I(\tau < 0)$,

where

$$G_* = G_0 + M'M,\ g_* = G_*^{-1}\left(G_0 g_0 + M' u\right),\ G_0 = \begin{bmatrix} C_0 & \mathbf{0}_{(2\times 2)} \\ \mathbf{0}_{(2\times 2)} & \Lambda_0 \end{bmatrix},\ g_0 = [d_0' \quad \lambda_0']',$$

$$M' = \begin{bmatrix} \frac{1}{\sqrt{h_1}} & \frac{1}{\sqrt{h_2}} & \cdots & \frac{1}{\sqrt{h_T}} \\ \frac{y_0}{\sqrt{h_1}} & \frac{y_1}{\sqrt{h_2}} & \cdots & \frac{y_{T-1}}{\sqrt{h_T}} \\ 1 & 1 & \cdots & 1 \\ I(S_1 = 2) & I(S_2 = 2) & \cdots & I(S_T = 2) \end{bmatrix}, u = \begin{bmatrix} \frac{y_1}{\sqrt{h_1}} \\ \frac{y_2}{\sqrt{h_1}} \\ \vdots \\ \frac{y_T}{\sqrt{h_T}} \end{bmatrix};$$

2. $p\left(\beta | \theta_{-\beta}, h, S, y\right) \propto f_N^{(2)}\left(\beta | b_*, \sigma^2 A_*^{-1}\right) I(|\varphi| < 0)$,

where

$$b_* = A_*^{-1}\left(\sigma^2 A_0 b_0 + W' \ln h\right),\ A_* = \sigma^2 A_0 + W'W,\ \ln h = \left(\ln h_1, \ln h_2, \ldots, \ln h_T\right)',$$

and

$$W' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \ln h_0 & \ln h_0 & \cdots & \ln h_{T-1} \end{bmatrix};$$

3. $p\left(\sigma^2 | \theta_{-\sigma^2}, h, S, y\right) \propto f_{IG}\left(\sigma^2 | \nu_1^*, \nu_2^*\right)$,

where

$$\nu_1^* = \frac{T}{2} + \nu_1, \quad \nu_2^* = \left[\frac{1}{\nu_2} + 0.5 \sum_{t=1}^{T} \left(\ln h_t - \mu - \varphi \ln h_{t-1}\right)^2\right]^{-1};$$

4. $p\left(p_{ii} | \theta_{-p_{ii}}, h, S, y\right) \propto f_B\left(p_{ii} | a_i^*, b_i^*\right),\ i = 1, 2$,

where

$$a_1^* = a_1 + n_{11}, \quad b_1^* = b_1 + n_{12}, \quad a_2^* = a_2 + n_{22}, \quad b_2^* = b_2 + n_{21},$$

and

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

Sampling the latent variables comprising h and S is more demanding and requires introduction of two additional steps within the Gibbs chain, both of which are presented below in some detail.

Sampling conditional variances

As noted by Jacquier, Polson, Rossi (1994) and Pajor (2003) in the case of a basic SV model, drawing each conditional variance h_t, $t = 1, 2, \ldots, T$, according to its non-typical conditional posterior, can be managed with the use of the Metropolis-Hastings (M-H) algorithm. Their approach is easily adapted in our setting.
In the SV-MS-M model we obtain the following expression for the kernel of the conditional posterior of h_t:

$$\begin{aligned} p\left(h_t|h_{-t}, S, \theta, y\right) &= p\left(h_t|h_{t-1}, h_{t+1}, S, \theta, y\right) \\ &\propto p\left(y_t|y_{t-1}, h_t, S_t, \theta\right) p\left(h_t|h_{t-1}, \theta\right) p\left(h_{t+1}|h_t, \theta\right) \\ &= \frac{1}{h_t^{1.5}} \exp\left\{-\frac{\left(y_t - m_t - \gamma_{S_t}\sqrt{h_t}\right)^2}{2h_t}\right\} \exp\left\{-\frac{(\ln h_t - \mu - \varphi \ln h_{t-1})^2}{2\sigma^2}\right\} \cdot \\ &\quad \cdot \exp\left\{-\frac{(\ln h_{t+1} - \mu - \varphi \ln h_t)^2}{2\sigma^2}\right\}, \end{aligned} \tag{24}$$

where h_{-t} denotes the vector h without its t-th covariate, and $m_t = \delta_0 - \delta_1 y_{t-1}$. For $t = T$ the third exponential factor in (24) disappears. The M-H step requires specifying an ancillary density, according to which candidate values of ht are generated within each cycle and then accepted with an appropriate probability; see Jacquier, Polson, Rossi (1994), Pajor 2003. We notice that under $\gamma_1 = \gamma_2 = 0$ the density in (24) collapses to the one obtained for a basic SV structure, in the case of which Jacquier, Polson, Rossi (1994) and Pajor (2003) use an Inverse-Gamma proposal density (with the only difference between the cited works being that the former authors assume $m_t = 0$, whereas the latter extends the observations to follow a first order autoregression, i.e. $m_t = \delta_0 - \delta_1 y_{t-1}$). The result also holds for the SV-M structure, in the case of which the conditional posterior of h_t assumes the form of (24) with the restriction $\gamma_1 = \gamma_2 = \gamma$. Consequently, in each of the three model specifications to generate candidate draws for h_t the very same Inverse-Gamma density of Pajor (2003) is utilized. Since the acceptance rate fluctuates between 80% and 85%, we infer that the approach proves efficient in our applications.

Sampling state variables

As regards generating the state variables, S_t's, the multi-move sampler of Carter and Kohn (1994) and Chib (1996) is suitably applied. Although the method has been developed for models without a stochastic volatility component and in different contexts, it can easily be adapted to our framework as the conditional variances may be perceived as additional, yet given (at this stage) parameters. The procedure hinges on the following factorization of the conditional posterior distribution of S:

$$\begin{aligned} p\left(S|y,h,\theta\right) &= p\left(S_T|y,h,\theta\right) \prod_{t=T-1}^{1} p\left(S_t|y,S_{(t+1:T)},h,\theta\right) \propto \\ &\propto p\left(S_T|y,h,\theta\right) \prod_{t=T-1}^{1} \left[p\left(S_t|y_{(1:t)},h,\theta\right) p\left(S_{t+1}|S_t,\theta\right)\right], \end{aligned} \tag{25}$$

where generally $x_{(m:n)} = (x_m, x_{m+1}, \ldots, x_n)$, $1 \leq m \leq n \leq T$, with the convention applying to both y and S. Following Chib (1996), in derivation of the second line in (25) we utilized the fact that

$$\begin{aligned} p\left(S_t|y,S_{(t+1:T)},h,\theta\right) &\propto p\left(S_t|y_{(1:t)},h,\theta\right) p\left(y_{(t+1:T)},S_{(t+1:T)}|y_{(1:t)},S_t,h,\theta\right) \\ &\propto p\left(S_t|y_{(1:t)},h,\theta\right) p\left(S_{t+1}|S_t,\theta\right) \cdot \\ &\quad \cdot p\left(y_{(t+1:T)},S_{(t+2:T)}|y_{(1:t)},S_t,S_{t+1},h,\theta\right) \\ &\propto p\left(S_t|y_{(1:t)},h,\theta\right) p\left(S_{t+1}|S_t,\theta\right), \end{aligned}$$

noting that the mass function $p\left(S_{t+1}|S_t,\theta\right) = p_{S_t,S_{t+1}}$ is independent of y and h and the term $p\left(y_{(t+1:T)},S_{(t+2:T)}|y_{(1:t)},S_t,S_{t+1},h,\theta\right)$ is independent of S_{t+1}.
Analogous to a general state-space model, Carter and Kohn (1994) and Chib (1996) construct a discrete filter for the evaluation of $p\left(S_t|y_{(1:t)},h,\theta\right)$. Desired samples from $p\left(S|y,h,\theta\right)$ are generated using the forward-filtering-backward-sampling scheme, consisting of the following steps (see Chib 1996).

1. *Forward-filtering step*

   (a) *Prediction step*: Determination of $p\left(S_t|y_{(1:t-1)},h,\theta\right)$, $t = 1,2,\ldots,T$.
   Assume that $p\left(S_{t-1}|y_{(1:t-1)},h,\theta\right)$ is already available (starting with $p(S_0)$ for $t = 1$). By the law of total probability we obtain:
   $$\begin{aligned} p\left(S_t|y_{(1:t-1)},h,\theta\right) &= \sum_{k=1}^{2} p\left(S_t|S_{t-1}=k,y_{(1:t-1)},\theta\right) \cdot p\left(S_{t-1}=k|y_{(1:t-1)},h,\theta\right) \\ &= \sum_{k=1}^{2} p\left(S_t|S_{t-1}=k,\theta\right) \cdot p\left(S_{t-1}=k|y_{(1:t-1)},h,\theta\right). \end{aligned}$$

   (b) *Update step*: Determination of $p\left(S_t|y_{(1:t)},h,\theta\right)$, $t = 1,2,\ldots,T$.
   By Bayes theorem we have:
   $$p\left(S_t|y_{(1:t)},h,\theta\right) = \frac{p\left(y_t,S_t|y_{(1:t-1)},h,\theta\right)}{p\left(y_t,S_t=1|y_{(1:t-1)},h,\theta\right) + p\left(y_t,S_t=2|y_{(1:t-1)},h,\theta\right)},$$

where

$$p\left(y_t, S_t | y_{(1:t-1)}, h, \theta\right) = p\left(y_t | y_{(1:t-1)}, S_t, h, \theta\right) p\left(S_t | y_{(1:t-1)}, h, \theta\right),$$

and

$$p\left(y_t | y_{(1:t-1)}, S_t, h, \theta\right) = f_N\left(y_t | \delta_0 + \delta_1 y_{t-1} + \gamma_{S_t}\sqrt{h_t}, h_t\right).$$

2. *Backward-sampling step*
Once the mass functions $p\left(S_t | y_{(1:t)}, h, \theta\right)$ are calculated (through a recursive run of the prediction and update steps):

   (a) Sample S_T from $p\left(S_T | y, h, \theta\right)$.

   (b) Sample S_t from

$$p\left(S_t | y, S_{(t+1:T)}, h, \theta\right) = \frac{p\left(S_t | y_{(1:t)}, h, \theta\right) p\left(S_{t+1} | S_t, \theta\right)}{p\left(S_{t+1} | y_{(1:t)}, h, \theta\right)},$$

   for $t = T-1, T-2, \ldots, 1$, where $p\left(S_{t+1} | y_{(1:t)}, h, \theta\right)$ and $p\left(S_t | y_{(1:t)}, h, \theta\right)$ are obtained from the prediction and update step, respectively.

3.4 Model comparison

Since it is three different models that are under consideration, their comparison in terms of the in-sample fit is of particular interest. Within the Bayesian methodology one needs to calculate the value of the marginal data density (sometimes called 'marginal likelihood') for each model:

$$p\left(y | M_l\right) = \int_{\Omega_l} p\left(\omega_{(l)}, y | M_l\right) d\omega_{(l)} = \int_{\Omega_l} p\left(y | \omega_{(l)}, M_l\right) p\left(\omega_{(l)} | M_l\right) d\omega_{(l)}, \tag{26}$$

where M_l denotes the l-th ($l = 1, 2, 3$) model, all the parameters and latent variables of which are arrayed in $\omega_{(l)} \in \Omega_l$. The $\dim(\Omega_l)$-tuple integral in (26) is numerically evaluated via the technique of Newton and Raftery (1994), according to the formula:

$$\widehat{p}\left(y | M_l\right) = \left[\frac{1}{N}\sum_{q=M+1}^{M+N} \frac{1}{p\left(y | \omega_{(l)}^{(q)}, M_l\right)}\right]^{-1}, \tag{27}$$

where M is the number of the burnt-in passes, N - the number of drawings from the joint posterior, q - the index of a single pass of the sampling procedure ($q = 1, 2, \ldots, M, \ldots, M+N-1, M+N$), and $\omega_{(l)}^{(q)}$ - the outcome on $\omega_{(l)}$ from the q-th cycle. The method is straightforward and despite its immanent numerical

instability, it proved satisfactory in our applications.
A pairwise comparison of different model structures is carried out by means of Bayes factors, which are calculated as:

$$BF_{kl} = \frac{p\left(y|M_k\right)}{p\left(y|M_l\right)}, \tag{28}$$

where the quantities on the right-hand side are estimated with (27). Under equal prior odds, i.e. $p\left(M_l\right) = \frac{1}{3}$, of each model, BF_{kl} equals the posterior odds ratio of M_k against M_l.

3.5 Forecasting

Bayesian forecasting requires construction of a joint predictive distribution of future observations and, in the case of their presence, variables comprising the hidden processes. Considering the SV-MS-M model, let us presume we are interested in inference on the future rates of return, $y_f = \left(y_{T+1}, y_{T+2}, \ldots, y_{T+K}\right)'$, conditional variances, $h_f = \left(h_{T+1}, h_{T+2}, \ldots, h_{T+K}\right)'$, and states of the mechanism governing regime changes, $S_f = \left(S_{T+1}, S_{T+2}, \ldots, S_{T+K}\right)'$, with K being the maximal forecast horizon. Then, the density of the joint predictive distribution is given by the formula:

$$p\left(y_f, h_f, S_f|y\right) = \int_{\Omega} p\left(y_f, h_f, S_f|y, \omega\right) p\left(\omega|y\right)\, d\omega, \tag{29}$$

A pseudo-random sample from (29) is obtained quite easily due to a straightforward factorization of the conditional predictive density on the right-hand side of (29):

$$p\left(y_f, h_f, S_f|y, \omega\right) = \prod_{t=T+1}^{T+K} f_N\left(y_t|\delta_0 + \delta_1 y_{t-1} + \gamma_{S_t}\sqrt{h_t}, h_t\right) f_N\left(\ln h_t|\mu + \varphi \ln h_{t-1}, \sigma^2\right) p\left(S_t|S_{t-1}, \theta\right), \tag{30}$$

where $p\left(S_t|S_{t-1}, \theta\right) = p_{S_{t-1}, S_t}$. Once a sample $\left\{\left(y_f^{(q)}, h_f^{(q)}, S_f^{(q)}\right), q = M+1, \ldots, N\right\}$ from (29) is generated, inference on any measurable function of (y_f, h_f, S_f), such as future asset prices and conditional standard deviations, is easily made.

4 Empirical study

We analyse ten series of daily logarithmic rates of return, calculated as $y_t = 100 \ln\left(\frac{x_t}{x_{t-1}}\right)$, where x_t denotes the asset closing price at time $t = 1, 2, \ldots, T$. Following the studies of Fiszeder and Kwiatkowski (2005) and Pipień (2007), who found that analysing the excess rates of return as opposed to 'crude' returns bears

little impact on the results, we consider the series $\{y_t, t = 1, 2, \ldots, T\}$ rather than the one of excess returns.
The main objective of this study is to examine the Warsaw Stock Exchange index (WIG) along with its sectoral subindices in search of a possibly regime-switching risk premium effect. To this end, three models are estimated for each series: the basic SV, SV-M and SV-MS-M model (see Table 1). Whether a particular series displays either a constant or regime-changing in-mean effect amounts to formal Bayesian model comparison. Hence, once the datasets are given a brief account of, we proceed with the results on the in-sample performance of the competing structures.

4.1 Data sets

Table 2 contains a detailed list of the series under consideration, whereas in Table 3 basic descriptive statistics are reported. Due to different sample sizes to some of the series we refer as the short series and to the others as the long series (see Table 2).

Table 2: Descriptive statistics of the analysed time series of the daily log-returns

(Sub)Index		Sample time range		Sample size,	Series
Full name	Abbrev. name	Since	To	T	description
WIG-chemicals	WIG-chem	Sep. 25, 2008	Sep. 01, 2009	234	short series
WIG-developers	WIG-dev	Jun. 21, 2007		551	
WIG-oil&gas	WIG-oil	Jan. 04, 2006		917	
WIG-media	WIG-med	Jan. 05, 2005		1168	
WIG	WIG	Jan. 07, 1998		2922	long series
WIG-banking	WIG-bank				
WIG-construction	WIG-cons				
WIG-IT	WIG-IT				
WIG-food	WIG-food				
WIG-telecom	WIG-tel				

Different sample path lengths across the WIG subindices result from the differences in the commencement dates of their publication. The sample size of the WIG series has been tailored to other long series. Note that in each case the sample period includes the current global financial crisis, which is a deliberate attempt to examine performance of the models in the presence of market turbulences, especially in forecasting terms.

4.2 General results

In Table 4 we present the number of the burnt-in sampling passes, the total number of which rises in line with the complexity of the model. The cycles are followed by 500000

Table 3: List of the analysed time series

Index	Min	Max	Mean	Std. dev.	Skewness	Kurtosis
WIG-chem	-9.157	8.542	-0.045	2.597	-0.354	4.135
WIG-dev	-11.725	11.030	-0.147	2.718	-0.160	5.015
WIG-oil	-9.351	9.846	-0.057	2.190	-0.035	4.437
WIG-med	-10.738	6.687	0.012	1.725	-0.453	5.723
WIG	-9.974	7.893	0.032	1.550	-0.279	5.930
WIG-bank	-14.436	8.880	0.050	1.866	-0.133	7.062
WIG-cons	-9.930	8.203	0.038	1.633	-0.193	5.865
WIG-IT	-11.070	9.351	-0.003	2.125	-0.001	5.463
WIG-food	-9.828	9.429	0.020	1.484	-0.324	8.053
WIG-tel	-10.399	9.351	-0.007	2.222	0.036	4.586

draws from the joint posterior in each case. Convergence of the MCMC procedure has been monitored by means of the ergodic averages and standard deviations, much in the vein of Bauwens and Lubrano (1998) and Pajor (2003). Visual inspection of the CUMSUM plots (not presented in the paper) implies that in the case of the SV-MS-M models convergence is attained after about 250000 to 500000 burnt-in passes (depending on the series). For the simpler structures, i.e. SV and SV-M, the sampler converges to the joint posterior markedly faster (although for the media, banking and construction subindices estimation of the SV-M model still required relatively more transient passes; see Table 4).

Table 4: Number of the burnt-in cycles generated for the modelled time series

Index	burnt-in cycles		
	SV (M_1)	SV-M (M_2)	SV-MS-M (M_3)
WIG-chem	100,000	400,000	1,000,000
WIG-dev	100,000	400,000	1,000,000
WIG-oil	100,000	400,000	1,000,000
WIG-med	100,000	1,000,000	1,500,000
WIG	100,000	500,000	1,500,000
WIG-bank	100,000	1,000,000	1,000,000
WIG-cons	200,000	1,000,000	1,500,000
WIG-IT	100,000	400,000	1,500,000
WIG-food	100,000	600,000	1,500,000
WIG-tel	100,000	400,000	1,000,000

We begin with the results on model comparison. Relevant quantities, including decimal logarithms of the marginal data density values and Bayes factors (against the SV specification), are displayed in Table 5. Apparently, for each modelled time series

our Markov switching model outperforms the two other structures in terms of the in-sample fit. However, taking the numerical instability of the Newton and Raftery procedure into account, only for the WIG-oil&gas, WIG-banking and WIG-food indices the differences in the marginal likelihoods may be perceived meaningful. For these series the logs of Bayes factors calculated for the SV-MS-M model against the SV specification amounted to 7.09, 30.22 and 11.26, respectively. The results imply that the SV-MS-M model fits the three datasets better than the SV structure by about as much as 7, 30 and 11 orders of magnitude. For the two long series the discrepancies in data fit between the SV-M and the regime-switching model are even larger (since the previous appears a less adequate model in both cases), whereas in the case of the WIG-oil&gas index the SV-MS-M model surpasses the constant-risk-premium specification only marginally (i.e. within the margins of a numerical error of the Newton and Raftery estimator).

Table 5: Decimal logarithms of the marginal likelihoods and Bayes factors, and ranks of the models

Index	SV (M_1)		SV-M (M_2)			SV-MS-M (M_3)		
	$\log[p(y\|M_1)]$	Rank	$\log[p(y\|M_2)]$	$\log(BF_{21})$	Rank	$\log[p(y\|M_2)]$	$\log(BF_{31})$	Rank
WIG-chem	-236.26	3	-236.01	0.24	2	-235.64	0.61	1
WIG-dev	-552.75	2	-553.14	-0.39	3	-551.92	0.83	1
WIG-oil	-856.35	3	-850.71	5.64	2	-849.25	7.09	1
WIG-med	-951.76	2	-951.97	-0.21	3	-951.33	0.43	1
WIG	-2171.11	2	-2173.02	-1.91	3	-2170.56	0.55	1
WIG-bank	-2363.20	2	-2365.87	-2.68	3	-2332.98	30.22	1
WIG-cons	-2219.26	2	-2225.00	-5.74	3	-2219.16	0.10	1
WIG-IT	-2532.25	3	-2528.66	3.59	2	-2525.95	6.30	1
WIG-food	-1976.34	2	-1977.90	-1.56	3	-1965.08	11.26	1
WIG-tel	-2665.73	2	-2667.20	-1.46	3	-2662.59	3.14	1

We notice that the results for the oil and gas industry index (unlike the ones for the two other series in question) do not reflect themselves in the characteristics of the marginal posterior distributions of the model-specific parameters (see Table 6). Just like for any of the short series, the posterior means of the transition probabilities p_{ii} are located close to 0.5, with the posterior distributions of p_{ii} additionally displaying a relatively large dispersion. What is more, although in some cases the posterior means of the regime-switching risk premium parameters, i.e. γ_1 and γ_2, seem to differ from zero, the posteriors of these quantities are, again, much dispersed. In view of the above, the hypothesis of a regime-switching in-mean effect seems to be overturned by the data. However, we highlight the fact, that the above results correspond to the marginal rather than joint distributions of the model-specific parameters. It may be the case, that as long as the marginal results for the oil and gas sector do

not qualitatively differ from the ones obtained for other short series, it is the joint presence of the model-specific parameters that contributes to the superior fit of the SV-MS-M model to the WIG-oil&gas data.

Table 6: Posterior means (and standard deviations) of the model-specific parameters

Index	SV-M (M_2)	SV-MS-M (M_3)				Index	SV-M (M_2)	SV-MS-M (M_3)			
	γ	p_{11}	γ_1	p_{22}	γ_2		γ	p_{11}	γ_1	p_{22}	γ_2
WIG-	-0.256	0.546	0.045	0.543	-0.663	WIG-	0.005	0.253	1.333	0.884	-0.276
chem	(0.273)	(0.278)	(0.378)	(0.270)	(0.467)	bank	(0.056)	(0.072)	(0.218)	(0.043)	(0.091)
WIG-	-0.014	0.588	0.330	0.602	-0.360	WIG-	0.020	0.777	0.237	0.704	-0.321
dev	(0.167)	(0.289)	(0.296)	(0.289)	(0.289)	cons	(0.058)	(0.278)	(0.222)	(0.285)	(0.328)
WIG-	-0.101	0.437	0.212	0.573	-0.313	WIG-	-0.110	0.798	0.052	0.371	-1.291
oil	(0.131)	(0.268)	(0.340)	(0.284)	(0.240)	IT	(0.052)	(0.287)	(0.232)	(0.242)	(0.906)
WIG-	-0.172	0.588	0.015	0.435	-0.548	WIG-	-0.061	0.844	0.432	0.902	-0.309
med	(0.127)	(0.291)	(0.213)	(0.284)	(0.450)	food	(0.043)	(0.101)	(0.228)	(0.177)	(0.288)
WIG	-0.127	0.605	0.090	0.461	-0.584	WIG-	-0.065	0.444	0.469	0.580	-0.535
	(0.057)	(0.304)	(0.245)	(0.272)	(0.558)	-tel	(0.059)	(0.247)	(0.332)	(0.228)	(0.507)

The above remarks, concerning the posterior means and dispersion, also hold for most of the long series, with two exceptions, however: the banking and the food sectors. For these two datasets clear dominance of the SV-MS-M model in respect of the in-sample fit is corroborated by the posterior characteristics of the SV-MS-M parameters (see Table 6). For the banking sector, the first regime, however scarcely persistent (since $E(p_{11}|y, M_3) \approx 0.25$), corresponds with a relatively large positive and significant risk premium (as we obtained $E(\gamma_1|y, M_3) \approx 1.33$ and the posterior standard deviation $D(\gamma_1|y, M_3) \approx 0.22$). The second regime is related to a weak negative effect, for we have $E(\gamma_2|y, M_3) \approx -0.28$ and $D(\gamma_2|y, M_3) \approx 0.09$. In the case of the WIG-food data both regimes are fairly persistent, yet the state-dependent in-mean effects are markedly weaker (see Table 6). Note that the results remain in accordance with our primary line of reasoning outlined at the beginning. Namely, as long as there is no sign of a constant (i.e. not regime-switching) risk premium *per se* (posterior means of γ being close to zero), both indices reveal its switching pattern.
As regards posterior inferences on the constant in-mean effect, in eight out of ten analysed series the posterior mean of γ is negative. Nevertheless, owing to a relatively large dispersion featured by the posterior distribution of γ, the effect is hardly significant in all the cases. For the other two assets, i.e. the WIG-banking and WIG-construction indices, although $E(\gamma|y, M_2)$ is positive, hardly can it be perceived significant; see Table 6.
Since only for the banking and food industry indices the posterior inference seem to indicate switches in the risk-return relationship, it is just these two datasets with the

results for which we are concerned in the remainder of this section.

4.3 Results for the WIG-banking and WIG-food indices

In Figures 1a and 1b we present the prior and posterior densities of the transition probabilities and Markov switching in-mean parameters for the banking and food indices. Marginal posterior densities for the WIG-food index are of less regular shapes than the ones obtained for the banking sector. Specifically, parameters related to the first regime, i.e. p_{11} and γ_1, feature evident humps, whereas the ones corresponding to the other state - heavy left tails. Nevertheless, a clear distinction between the priors and posteriors in both cases of the analysed series implies that the data bears strong information on the switching nature of the risk premium.

Table 7: Posterior means (and standard deviations) of the common parameters

Parameter	WIG-banking			WIG-food		
	SV (M_1)	SV-M (M_2)	SV-MS-M (M_3)	SV (M_1)	SV-M (M_2)	SV-MS-M (M_3)
δ_0	0.055	0.050	0.154	0.051	0.101	0.117
	(0.026)	(0.076)	(0.086)	(0.017)	(0.040)	(0.054)
δ_1	0.087	0.087	0.062	0.072	0.071	0.015
	(0.019)	(0.019)	(0.024)	(0.020)	(0.020)	(0.032)
μ	0.020	0.020	0.014	0.010	0.010	0.006
	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)
φ	0.977	0.976	0.978	0.956	0.956	0.955
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.010)
σ^2	0.033	0.033	0.031	0.102	0.101	0.104
	(0.007)	(0.007)	(0.006)	(0.019)	(0.019)	(0.020)
σ^{-2}	31.641	31.415	33.807	10.137	10.255	10.001
	(6.720)	(6.623)	(6.875)	(1.914)	(1.956)	(1.976)

Since it is the time-varying volatility that underlies all the models under consideration, it is interesting whether the posterior results indicate any differences across the three specifications in terms of the volatility pattern and parameters. As regards posterior characteristics of the marginal densities of the volatility parameters, they seem robust to the model specification (see Table 7), perhaps with an exception of the intercept, μ, whose posterior mean is marginally lower in the switching model. We note, however, that extending a basic stochastic volatility structure to the SV-M and SV-MS-M models may exert some influence on the common conditional mean parameters. Regarding the switching specification, for both of the analysed series we observe a rise in the posterior mean of the intercept and a drop in the one of the autoregression coefficient. Intuitively speaking, the latter could stem from the

Figure 1a: Prior and posterior densities (dashed and solid lines, respectively) of the model-specific parameters in the SV-MS-M model: WIG-banking

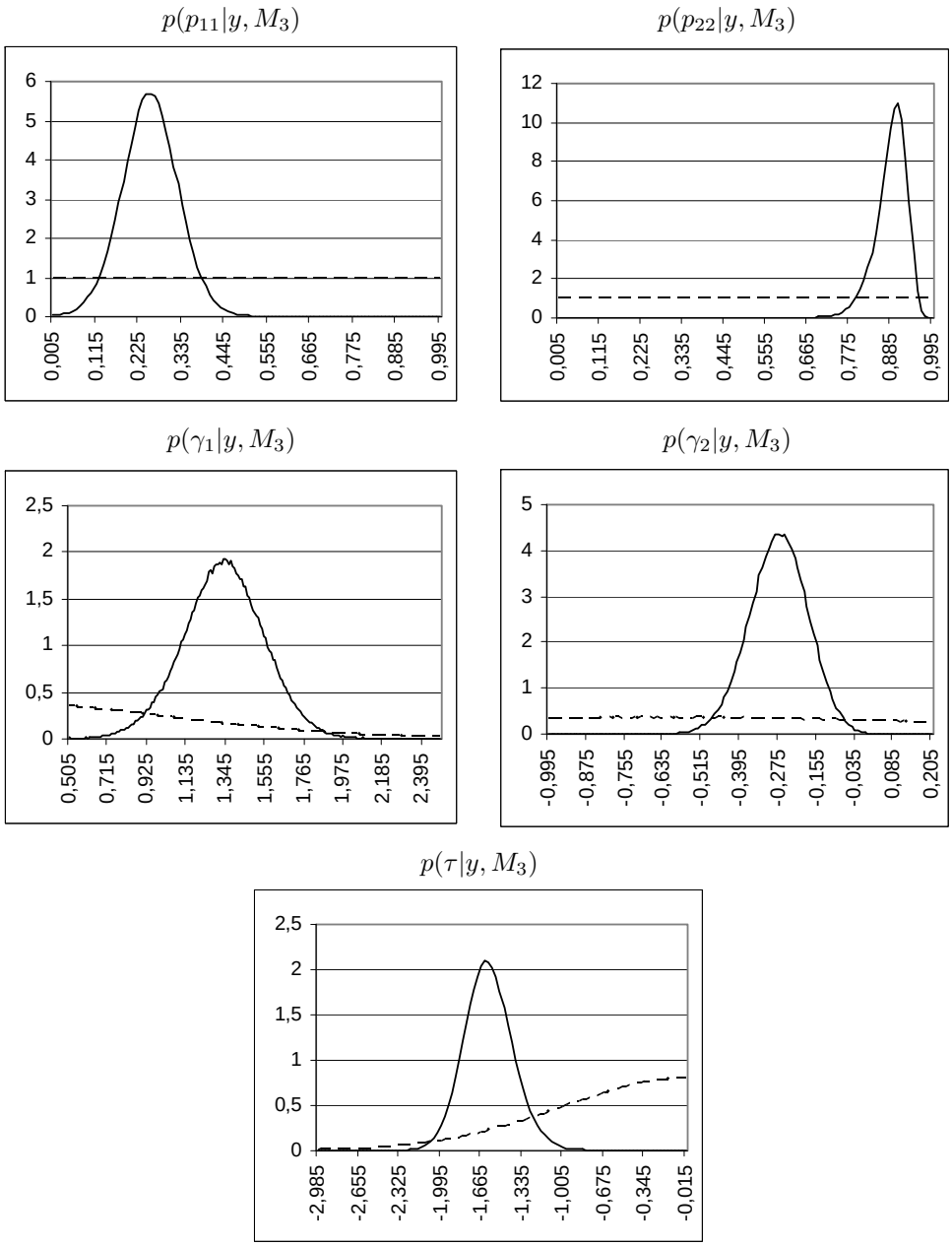
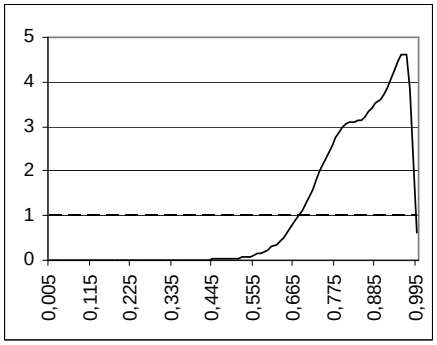
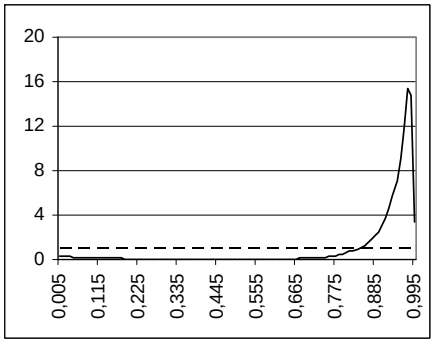


Figure 1b: Prior and posterior densities (dashed and solid lines, respectively) of the model-specific parameters in the SV-MS-M model: WIG-food

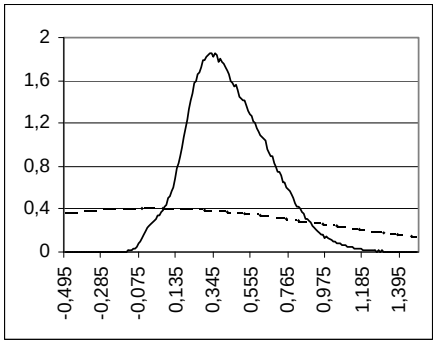
$p(p_{11}|y, M_3)$



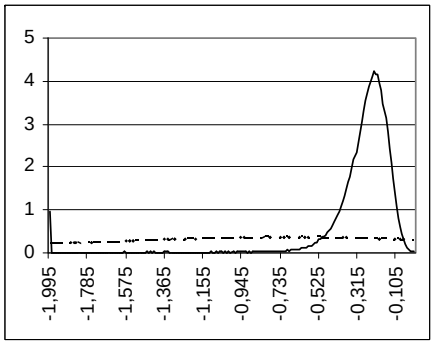
$p(p_{22}|y, M_3)$



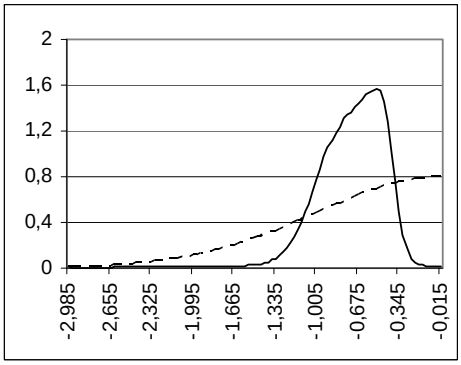
$p(\gamma_1|y, M_3)$



$p(\gamma_2|y, M_3)$

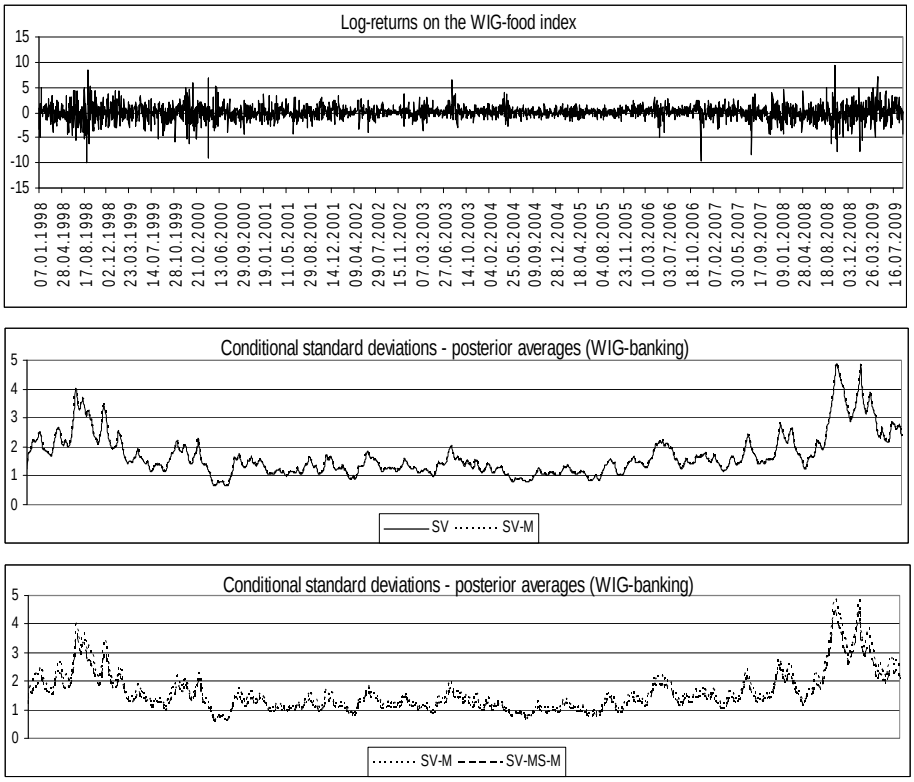


$p(\tau|y, M_3)$



inclusion of a hidden Markov chain that may capture a 'part' of the autocorrelation in the observable process, resulting in a lower posterior mean of the autoregression coefficient. The aforesaid robustness of the volatility parameters to the

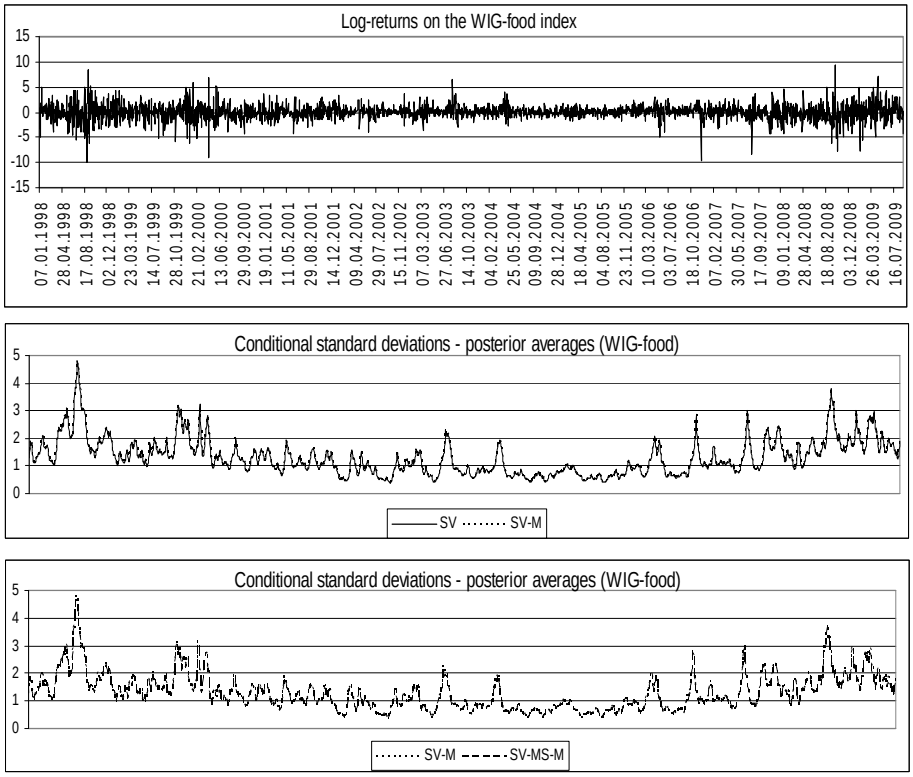
Figure 2: Posterior averages of the conditional standard deviations across the models - results for the WIG-banking index



specification of the observation equation, reflects itself in almost identical patterns (across the models) of the posterior averages of the latent conditional standard deviations (see Figures 2 and 3). Nonetheless, the ones obtained in the SV-MS-M model estimated for the banking sector are marginally, yet systematically lower than the ones in the two other specifications.

We proceed with the inference on the regimes. In view of the apparent differences between the prior and posterior densities of the ergodic probabilities, π_i, and expected durations, $Dur_i, i = 1, 2$ (see Figure 4), it is clear that the data bears significant information on the underlying Markov process (although the priors, especially for durations, seem fairly informative). The results for the banking sector, however, may

Figure 3: Posterior averages of the conditional standard deviations across the models - results for the WIG-food index



appear not much appealing. Since $E\left(Dur_1|y, M_3\right) \approx 1.35$ and $E\left(Dur_2|y, M_3\right) \approx 9.85$ (see Table 8) we gather that the state of a strong positive risk premium holds, on average, merely for one session day, whereas the other one (of a negative effect) - for ten days. The results are closely related to the in-sample pattern of the probabilities $\Pr(S_t = 1|y, M_3)$, which are evaluated at each data point $t = 1, 2, \ldots, T$, according to the formulae:

$$\Pr\left(S_t = 1|y, M_3\right) \approx \frac{1}{N} \sum_{q=M+1}^{M+N} I\left(S_t^{(q)} = 1\right), \tag{31}$$

and which may be perceived as 'smoothed' probabilities of the first regime; see Figure 5. The probabilities in question are very close to the probabilities of a positive in-mean effect, i.e. $\Pr\left(\gamma_{S_t} > 0|y, M_3\right), t = 1, 2, \ldots, T$, which stems from the fact, that probabilities $\Pr(\gamma_1 < 0|y, M_3)$ and $\Pr(\gamma_2 > 0|y, M_3)$ are negligible; see Figure 1a and Figure 1b.

We observe that positive risk premium is a very elusive and short-lived phenomenon in the banking industry. Usually, the first regime could be assigned only to single observations, which may question the very notion of a 'regime' as a period of some longer duration. Additionally, for the most part of the sample it is hard to distinguish between periods of a positive and negative in-mean effect (see the bottom plot in Figure 5). Due to the intense variability displayed by the switching mechanism, hardly can we think of any economic reasons behind its behaviour, especially in terms of the events that might have triggered the occurrences of a positive risk premium. Heuristically, one could think of an apparent influence of the foreign stock and currency markets on the Polish banking sector. However, no specific economic interrelations could be pointed as yet. On the other hand, we have found a statistical explanation for a somewhat erratic behaviour of the process governing the regime shifts. As noted in Section 2, our SV-MS-M specification may be represented as a simple stochastic volatility model with two-component Markov mixture errors, ξ_t's, whose distributions: unconditional and conditional upon the lagged state, may feature asymmetry. Therefore, posterior inference is made on the skewness coefficients, including $Sk\left(\xi_t\right) \equiv Sk$ and $Sk\left(\xi_t | S_{t-1} = i\right) \equiv Sk_i$, $i = 1, 2$ (see Table 9 and Figure 7). In the case of the WIG-banking series, a strong evidence of a positive skewness is found, with regard to both the unconditional and S_{t-1}-conditional distribution of ξ_t. Despite the fact that the priors of Sk and Sk_i (implied by the prior structure presented in Section 3.2) are markedly concentrated around zero and feature heavy tails (and thus appearing quite informative; see Figure 7), the data shifts the probability mass to the right, resulting in the skewness coefficients' posterior densities tightened around 0.2. It seems that our regime-switching model captures and implies conditional skewness of the modelled log-returns rather than discrete changes in the risk-return relationship themselves. Hence, we deem it worthwhile to compare in terms of the data fit two specifications: the SV-MS-M model and a basic SV structure with the error term following the skewed Normal or Student-t distribution. We shall address the issue elsewhere.

Table 8: Posterior means (and standard deviations) of the ergodic probabilities and expected durations

Index \ Quantity	π_1	Dur_1	π_2	Dur_2
WIG-banking	0.133	1.353	0.867	9.850
	(0.045)	(0.136)	(0.045)	(3.788)
WIG-food	0.310	12.971	0.690	31.399
	(0.186)	(47.311)	(0.186)	(32.345)

Figure 4: Prior and posterior densities (dashed and solid lines, respectively) of the ergodic probabilities and expected durations

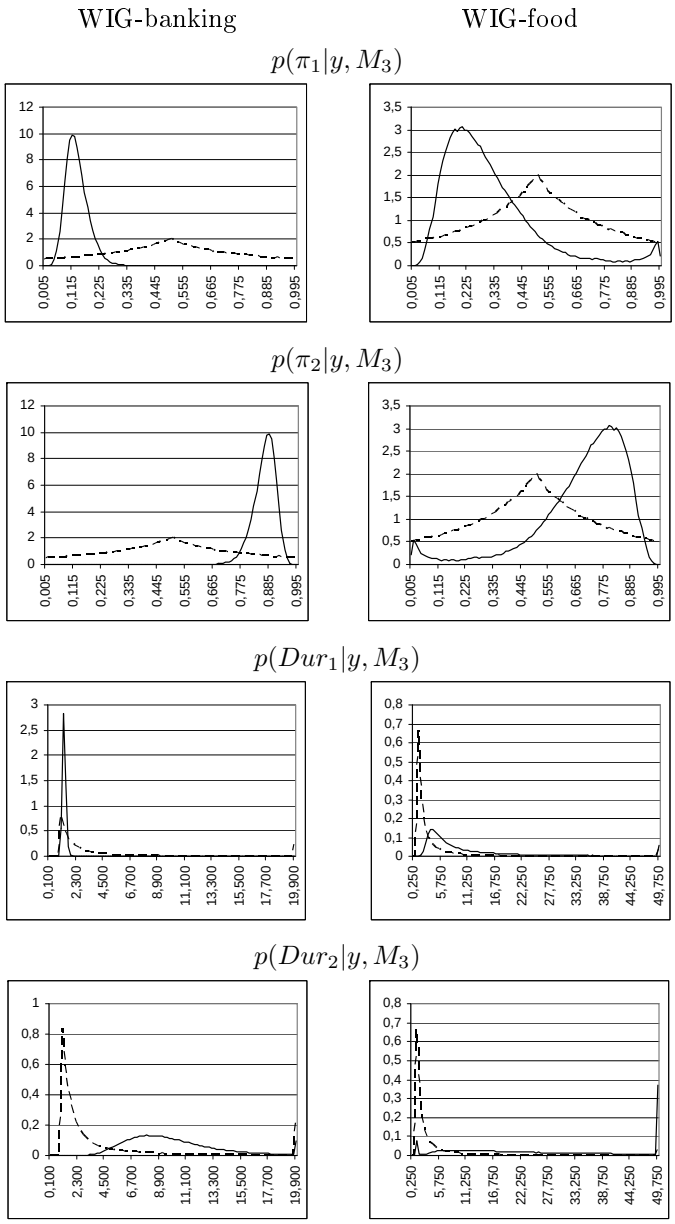
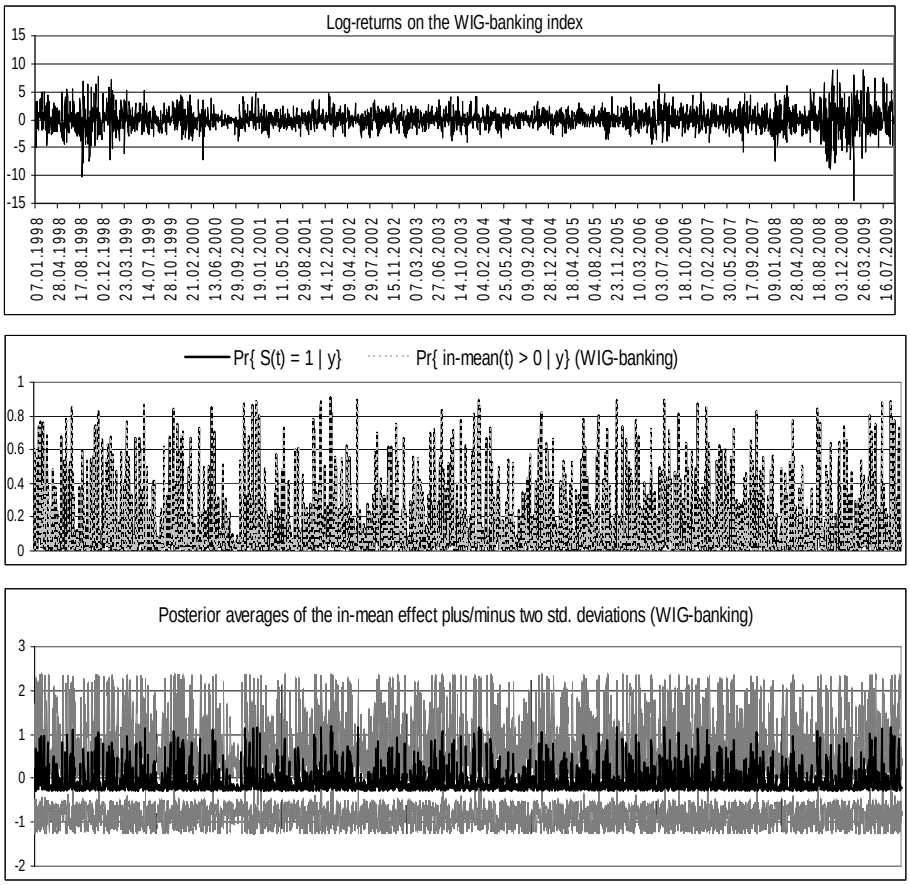


Figure 5: Posterior probabilities of the first regime and averages of the in-mean effect - results for the WIG-banking index



Note: By 'Pr{in-mean$(t) > 0|y$}' we denote $\Pr\left(\gamma_{S_t} > 0|y\right)$.

The results for the food industry index are distinct. As indicated by the posterior means of the transition probabilities: $E\left(p_{11}|y, M_3\right) \approx 0.84$ and $E\left(p_{22}|y, M_3\right) \approx 0.9$ (see Table 6), both regimes are fairly persistent. Although posterior averages of the expected durations are higher than the ones obtained for the WIG-banking series, huge posterior dispersion featured by the posteriors of Dur_1 and Dur_2, as well as irregularities in their densities' shape (possibly resulting from the data clashing with quite informative priors) preclude a precise inference on the lasting of the regimes (see Table 8 and Figure 4). Incidentally, we note that as large as the posterior standard deviation of Dur_1 may appear (being equal to 47.311), it is sensitive to

Figure 6: Posterior probabilities of the first regime and averages of the in-mean effect - results for the WIG-food index

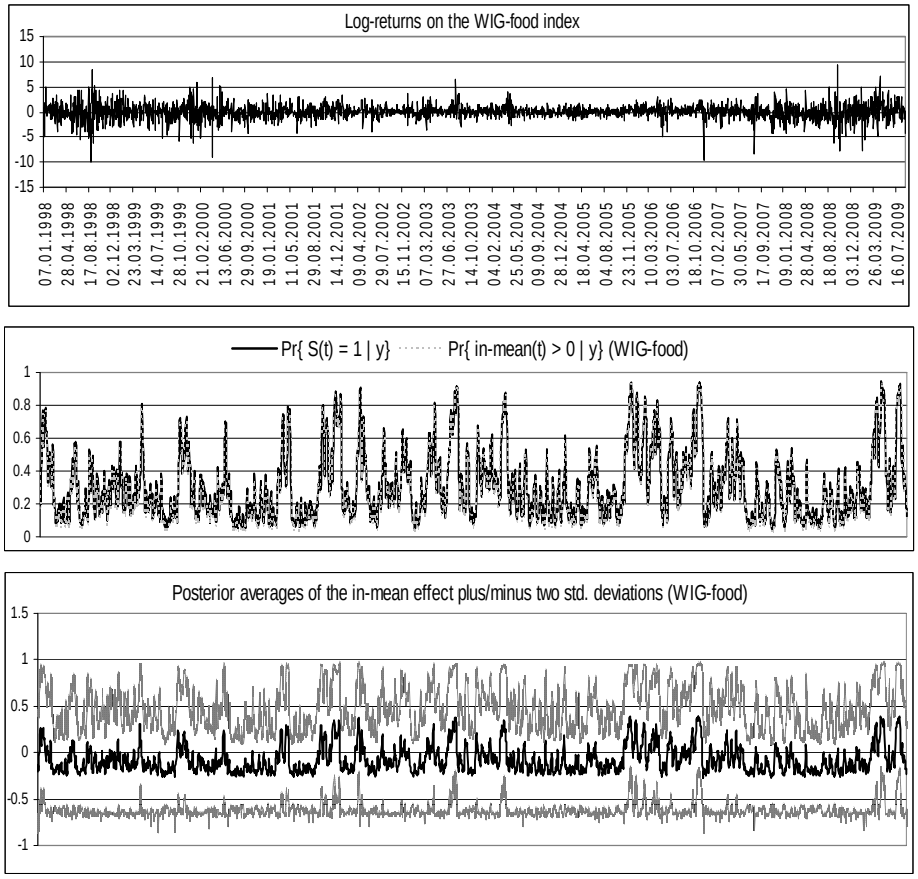


Table 9: Posterior means (and standard deviations) of the skewness coefficients of the error term ξ_t in the SV-MS-M model

Index \ Quantity	Sk	Sk_1	Sk_2
WIG-banking	0.230	0.206	0.220
	(0.059)	(0.071)	(0.056)
WIG-food	0.100	-0.109	0.069
	(0.122)	(0.093)	(0.104)

single, rare and close-to-unit MCMC realizations of the relevant transition probability, p_{11}. For example, if the two largest elements of the pseudo-random sample from

$p\left(p_{11}|y, M_3\right)$, being equal to 0.99995 and 0.99994, were to be discarded, then the posterior standard deviation of Dur_1 would drop to 29.849 (with $E\left(Dur_1|y, M_3\right) \approx$ 12.898 being only slightly modified as compared with its current, full sample value of 12.971; see Table 8). However, the issue does not concern the posterior of the other state's duration, as the MCMC chain does not venture into similar regions of the parameter space along the p_{22} coordinate.
Similarly as in the case of the banking index, probabilities of a positive risk premium are very close to the ones corresponding with the first regime (see Figure 6), much due to the same reason as before. Owing to a greater persistence of the underlying Markov chain, the latter reveals a more regular pattern as compared with the WIG-banking series, although still for the most part of the sample classifying the observations into the two regimes entails much uncertainty. Nevertheless, short episodes of a positive in-mean effect, lasting for a dozen of session days or so, may be discerned. As regards the inference on the skewness coefficients, Sk and Sk_i, their posterior densities are closely tightened near zero (see Figure 7 and Table 9), which indicates an almost symmetric distribution of the error term ξ_t (both the unconditional and S_{t-1}-conditional one), and, hence, a symmetry of the conditional distribution of the data.

Table 10: Predictive probabilities of the first regime, $\Pr(S_{T+k} = 1|y, M_3)$

Index	Forecast horizon									
	1	2	3	4	5	6	7	8	9	10
WIG-banking	0.1204	0.1308	0.1332	0.1331	0.1334	0.1330	0.1329	0.1327	0.1333	0.1332
WIG-food	0.1838	0.2138	0.2356	0.2496	0.2609	0.2699	0.2768	0.28217	0.2862	0.2902

Lastly, we report on the forecasting performance of the considered models. For both series we build joint predictive distributions of future logarithmic rates of return, y_{T+k}, and asset prices, x_{T+k}, over ten subsequent session days within the period September 2 - 15, 2009 ($k = 1, 2, \ldots, K = 10$). Also, predictive probabilities of the first regime, i.e. $\Pr(S_{T+k} = 1|y, M_3)$, are of interest. In Figures 8 and 9 predictive quantiles of the future log-returns and asset prices are plotted. It seems that extending a simple SV model to the ones featuring either a constant or regime-switching in-mean parameter, makes little contribution to the prediction. With regard to both analysed time series, visually, no significant differences can be spotted between the corresponding quantiles obtained from different model structures. Additionally, and perhaps to one's dismay, we note that in each model the 0.9-prediction intervals (i.e. the ones bounded by the 0.05- and 0.95-quantile) fail to encompass the very first realization of both the price and the return (their observed values fall between the 0.01- and 0.02-quantile, regardless of the model specification). Similarly, in the case of the food industry index the fourth realized value of the log-return falls beyond the 0.95-quantile (the observed value falls between the 0.96- and 0.97-quantile, regardless of the model specification).

Figure 7: Prior and posterior densities (dashed and solid lines, respectively) of the skewness coefficients of ξ_t in the SV-MS-M model

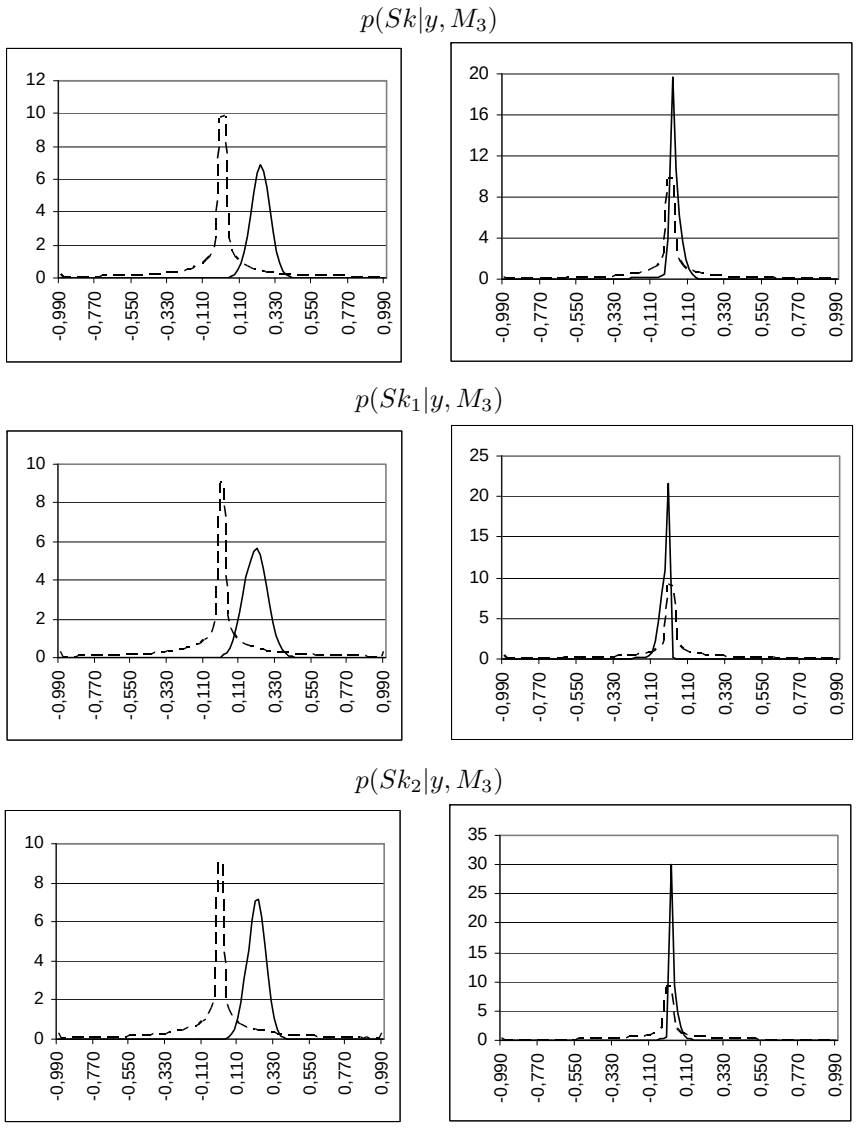
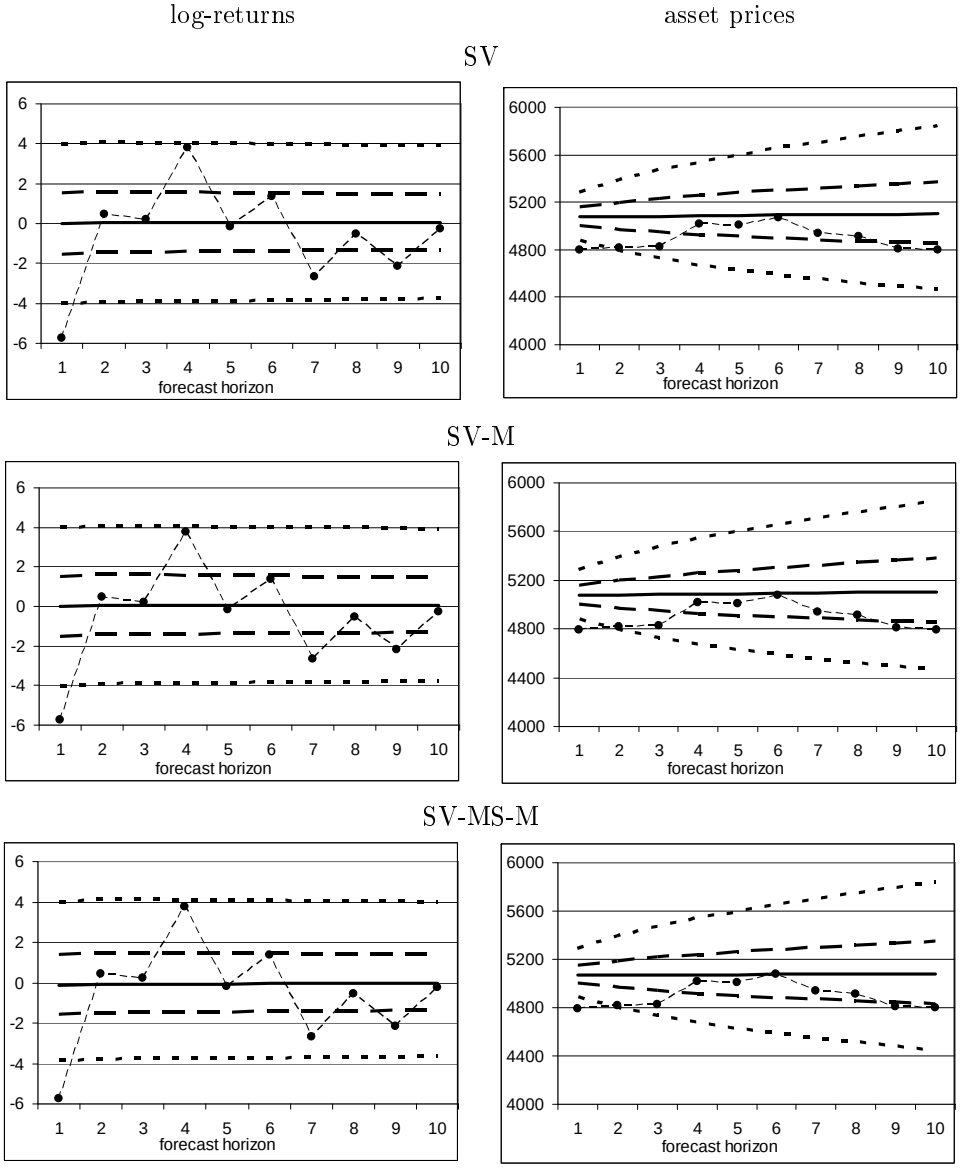
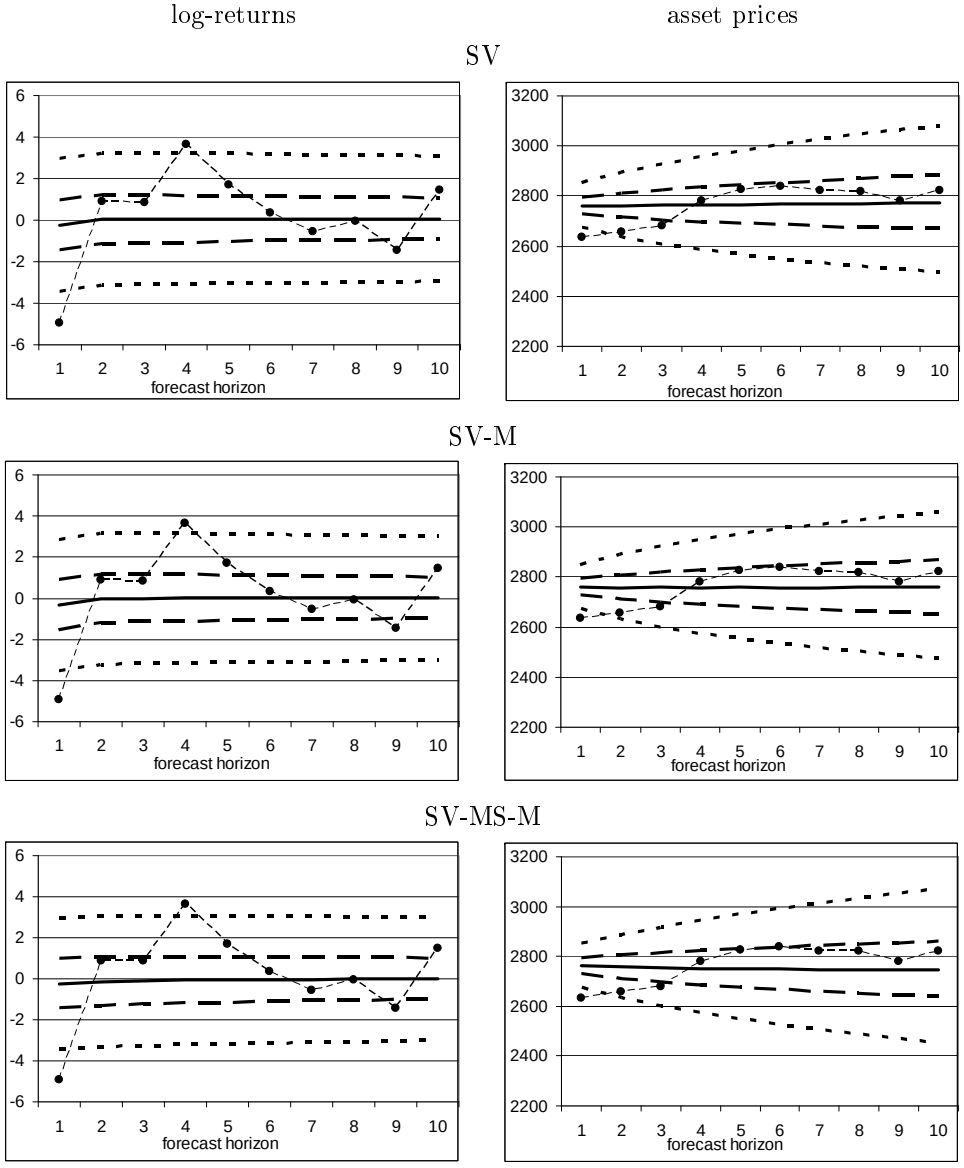


Figure 8: Predictive quantiles of future log-returns and asset prices - results for the WIG-banking index



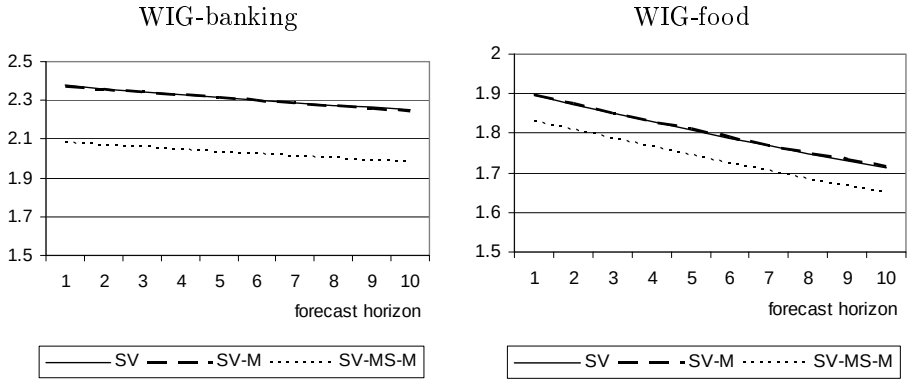
Note: In each plot the upper (bottom) dotted line represents 0.95-quantile (0.05-quantile), the upper (bottom) dashed line - 0.75-quantile (0.25-quantile), the solid line - median; the observed values' trajectory is plotted with dot markers linked with a thin dashed line.

Figure 9: Predictive quantiles of future log-returns and asset prices - results for the WIG-food index



With regard to both indices, we note that unlike the price quantiles, which tend to diverge with the forecast horizon, dispersion of the marginal distributions of the future returns changes little. In fact, it slightly decreases throughout the prediction period. The latter - as counterintuitive as it may appear - can be attributed to a relative drop in the asset price volatility noticed at the end of the sample period (see Figures 2 and 3), which causes the volatility forecasts to indicate further decline in the market uncertainty (see Figure 10).
As far as the predictive probabilities of the first regime are concerned, although we note their relative rise with the forecast horizon, chances of a positive risk premium remain quite low throughout (approximately 0.133 at the final prediction day for the WIG-banking, and 0.290 for the WIG-food index; see Table 10).

Figure 10: Predictive means of the conditional standard deviations



5 Concluding remarks

In the paper we proposed a new univariate stochastic volatility model incorporating a relationship between contemporaneous rates of return and conditional volatility, and, simultaneously, allowing for its possibly regime-changing pattern. The switching mechanism is governed by a two-state homogenous and ergodic Markov chain. Along with a simple SV and SV-in-Mean structures, the new specification is treated within the Bayesian methodology, exploiting MCMC simulation methods to allow posterior analysis of all the unknown quantities of the model.
Among ten analysed Polish stock market series of daily logarithmic growth rates, only two show clear evidence of a switching volatility-in-mean effect: the WIG-banking and WIG-food indices. For both of the series, the first regime corresponds with a positive risk premium. In the case of the banking index the effect is markedly stronger, although extremely short-lived. Contrary results with respect to the magnitude and

duration of a positive risk premium have been obtained for the other index. In both considered cases, however, the second state, which is related to a negative in-mean effect, seems to prevail. Additionally, for the banking index indications of a positively skewed conditional distribution of the returns have been detected and proposed as a statistical explanation for the superior in-sample fit of the SV-MS-M model. In terms of forecasting, all three models appear to perform comparably for both assets.
A possible reason for which the evidence of regime-switching risk premium among the analysed series is fairly scarce may reside in the fact that the stock market indices are some averages of individual stock prices. Thus, the effect, which could be featured by only certain shares, may be subdued once the stock aggregates are under consideration. It follows that investigating switches in the risk premium may be more justifiable in the case of individual stock prices rather than market indices.
Apart from shifting attention from stock indices to single share prices, further research should also address the issue of the number of the underlying Markov chain states as well as Bayesian model comparison of the SV-MS-M model with a SV structure featuring the skewed Normal or Student-t error term in the observation equation.

Acknowledgements

Research supported by a grant from Cracow University of Economics.
The author would like to thank Jacek Osiewalski for stimulating discussions and apt remarks that helped to significantly improve the rigour of the paper. Anna Pajor's and an anonymous referee's perceptive comments and insights are greatly acknowledged.

References

[1] Backus D.K., Gregory A.W., (1993), Theoretical Relations Between Risk Premiums and Conditional Variances, *Journal of Business & Economic Statistics* 11 (no. 2), 177-185.

[2] Bauwens L., Lubrano M., (1998), Bayesian inference on GARCH models using the Gibbs sampler, *Econometrics Journal* 1, C23-C46.

[3] Carter C.K., Kohn R., (1994), On Gibbs sampling for state space models, *Biometrika* 81, 541-553.

[4] 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

[5] 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.

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

[7] Chou R., Engle R.F., Kane A., (1992), Measuring Risk Aversion from Excess Returns on a Stock Index, *Journal of Econometrics* 52, 201-224.

[8] Engle R.F., (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987-1007.

[9] Engle R.F., Lilien D.M., Robins R.P., (1987), Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* 55, 525-542.

[10] Fiszeder P., Kwiatkowski J., (2005a), Model GARCH-M ze zmiennym parametrem - analiza wybranych spółek i indeksów notowanych na GPW w Warszawie (GARCH-M Model with Time-Varying Parameter - Analysis of Selected Stock and Indices Quoted on the WSE, in Polish), *Statistical Review (Przegląd Statystyczny)* 52 (no. 3), 73-88.

[11] Fiszeder P., Kwiatkowski J., (2005b), Dynamic Analysis of Relation between Expected Return and Conditional Variance, *Acta Universitatis Nicolai Copernici* 372, 85-98.

[12] Gärtner D., (2007), Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, *Working Paper* No. 0719, Socioeconomic Institute, University of Zürich.

[13] Glosten L.R., Jagannathan R., Runkle D.E., (1993), On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* 48, 1791-1801.

[14] Hamilton J. D., (1989), A New approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357-384.

[15] Hwang S., Satchell S.E., Pereira P.L.V., (2004), Stochastic Volatility Models with Markov Regime Switching State Equations, *Journal of Business and Economic Statistics* 16,

[16] Jacquier E., Polson N., Rossi P., (1994), Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economic Statistics* 12, 371-417.

[17] Kalimipalli M., Susmel R., (2001), Regime-switching stochastic volatility and short-term interest rates, *CEMA Working Papers*, available at: `http://ideas.repec.org/p/cem/doctra/197.html`

[18] Koopman S.J., Hol Uspensky E., (2002), The Stochastic Volatility In Mean Model: Empirical Evidence from International Stock Markets, *Journal of Applied Econometrics* 17, 667-689.

[19] Kwiatkowski Ł., (2009a), Markov switching SV processes in modelling volatility of financial time series, *Statistical Review (Przegląd Statystyczny)*, 56 (1), 147-168.

[20] Kwiatkowski Ł., (2009b), Markov switching in stochastic variance. Bayesian comparison of two simple models, *Folia Oeconomica Cracoviensia* 49-50, 107-141.

[21] N'dri K.L., (2008), An Empirical Study of the Relation between Stock Market Returns and Volatility in the BRVM, *International Research Journal of Finance and Economics* 14, 8-14.

[22] Nelson D.B., (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.

[23] Newton M.A., Raftery A.E., (1994), Approximate Bayesian inference by Weighted Likelihood Bootstrap (with discussion), *Journal of the Royal Statistical Society, series B* 56, 3-48.

[24] Osiewalski J., Pipień M., (2000), GARCH-In-Mean through skewed t conditional distributions: Bayesian inference for exchange rates, [in:] *26-th International Conference MACROMODELS'99*, [ed:] W. Welfe, P. Wdowiński, Łódź, 354-369.

[25] Pajor A., (2003), Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych (Stochastic Volatility Processes in Bayesian Analysis of Financial Time Series, in Polish), Cracow University of Economics.

[26] Pipień M., (2007), An approach to measuring the relation between risk and return. Bayesian analysis for WIG data, *Folia Oeconomica Cracoviensia* 48, 95-117.

[27] Pipień M., Osiewalski J., (2001), Model GARCH-M(1,1): specyfikacja i estymacja bayesowska (A GARCH-M(1,1) Model: Specification and Bayesian Estimation, in Polish), *Prace Naukowe Akademii Ekonomicznej we Wrocławiu* (no. 895), 188-197.

[28] Scruggs J.T., (1998), Resolving the Puzzling Intertemporal Relation between the Market Risk Premium and Conditional Market Variance: A Two-Factor Approach, *Journal of Finance* 53, 575-603.

[29] Shibata M., Watanabe T., (2005), Bayesian analysis of a Markov switching stochastic volatility model, *Journal of Japan Statistical Society* 35, 205-219.

[30] Smith D. R., (2002), Markov-Switching and Stochastic Volatility Diffusion Models for Short-Term Interest Rates, *Journal of Business and Economic Statistics* 20, 183-197.

[31] So M. K. P., Lam K., Li W. K., (1998), A stochastic volatility model with Markov switching, *Journal of Business and Economic Statistics* 16, 244-253.

[32] Valls Pereira P.L., (2004), How Persistent is Volatility? An Answer with Stochastic Volatility Models with Markov Regime Switching State Equations, *Finance Lab Working Papers* flwp_59, Finance Lab, Ibmec São Paulo.