Journal of the Nigerian Association of Mathematical Physics
Volume 47 (*July, 2018 Issue*), **pp19 – 28**
© *J. of NAMP*

REGIME-SWITCHING STOCHASTIC VOLATILITY MODELS FOR FORECASTING STOCK MARKET PRICES

Onyeka-Ubaka J. N. and Ogundeji R. K.

Department of Mathematics, University of Lagos, Nigeria

Abstract

Nonlinearity plays an important role in forecasting economic and financial variables. The paper proposes a regime-switching stochastic volatility (RSV) model with three regimes (negative jump, normal price, positive jump (spike)) where the transition matrix depends on explanatory variables. Gibbs Sampling-based Markov Chain Monte Carlo algorithm is used to estimate parameters of the models. Daily stock prices from the banking sector (First bank of Nigeria (FBN), United Bank for Africa (UBA), Guaranty Trust Bank (GTB) and Zenith Bank (ZEB) obtained from Nigeria Stock Exchange (NSE) are used to fit the model parameters, using Maximum Likelihood Estimation. The fit of the regime-switching model to the data is compared with the generalized autoregressive conditionally heteroskedastic (GARCH) family of models. The results show that RSV facilitates better capturing important features of the (joint) dynamics of the stock and volatility and is able to consistently match the selected stock market prices. In the forecast performance, the proposed model (RSV) outperforms other GARCH family of models in the in-sample and out-of-sample results based on our model diagnostics.

Key words: Forecasting, Markov switching, Stochastic Volatility, Maximum Likelihood Estimation

1.0 Introduction

Stochastic volatility (SV) models are increasingly important because they capture a richer set of empirical and theoretical characteristics than other volatility models. First, stochastic volatility models generate return distributions similar to what is empirically observed: the return distribution has a fatter left tail and kurtosis compared to normal distributions, with tail asymmetry controlled by the correlation between the stock and the volatility process [1-3]. Second, stochastic volatility models allow reproduction of the main features of the volatility behaviour: mean reversion and volatility clustering [4-5]. Third, stochastic volatility with a zero correlation always produces implied volatilities with a smile [6]. Fourth, Trolle and Schwartz [7] developed a tractable and flexible stochastic volatility multifactor model of the interest rates term structure. This model allows them to match the implied cap skews and the dynamics of implied volatilities. The "spiky" character of stock market prices suggests that there exists a nonlinear switching mechanism between normal and low/high states or regimes. The requirement of stochastic jump arrival probabilities directly leads to regime switching models. Markov regime-switching (MS) models seem to be a natural candidate for modeling such nonlinear and complex structure [8-10]. The rationale behind the regime-switching framework is that the market may switch from time to time between, say, a stable low-volatility state and a more unstable high-volatility regime. Periods of high volatility may arise, for example, because of short-term political or economic uncertainty.

This paper is organized as follows: Section 2 describes Regime-Switching Stochastic Volatility (RSV) model. Section 3 focuses on Methodology. Section 4 deals with Results and Materials and Section 5 present Conclusion.

2 Regime-Switching Stochastic Volatility (RSV) model

In this section, we provide the details on regime-switching stochastic volatility model for stock market prices. We first introduce the stochastic volatility model that exists within each of the three regimes (negative jump, regular, positive jump), followed by a description of the transition dynamics between the regimes. Under the regime-switching stochastic volatility model, it is assumed that the stock return process lies in one of K regimes or states. Let s_t denote the regime applying in the interval [t, t+1) (in weeks), $s_t = 1,\ 2,\ \cdots,\ K$ and r_t be the total return index value at t; then

$$\log \frac{r_{t+1}}{r_t}\Big| s_t \sim \mathrm{N}\left(\mu_{s_y},\ \sigma^2_{s_t}\right) \qquad (1)$$

Corresponding Author: Onyeka-Ubaka J.n., Email: jonyeka-ubaka@unilag.edu.ng, Tel: +2348059839937, +2348033528911(ORK)

In the RSV specification, the drift term of the conditional variance is a function of both current and previous period states. Consider the short-term stock return process where we describe an SV model that incorporates regime-switching as:

$$r_t - r_{t-1} = a_0 + a_1 r_{t-1} + \sigma_t r_{t-1}^{2\alpha} \varepsilon_t, \quad \alpha = 0.5$$

$$\left(\ln(\sigma_t^2) - \mu_{s_t}\right) = \phi_1\left(\ln(\sigma_{t-1}^2) - \mu_{s_{t-1}}\right) + \sqrt{\sigma_\eta^2}\,\eta_{t-1}$$

$$\mu_{s_t} = \beta + \gamma s_t \quad \gamma > 0 \qquad s_t = \{1,\ 2,\ \cdots,\ k\}$$

$$\mathrm{P}\left[s_t = j \middle| s_{t-1} = i\right] = p_{ij} \qquad \text{with } i = 0,\ 1 \text{ and } j = 0,\ 1 \tag{2}$$

Model (2) is referred to as a Regime-switching Stochastic Volatility (RSV) model where r_t is the short-term stock return, σ_t^2 is the conditional variance of the stock return, *a* captures the levels effect in the model, μ_{s_t} is the stationary mean of the natural log of σ_t^2, ϕ_1 measures the degree of persistence of $\ln(\sigma_t^2)$, ε_t and η_t represent shocks to the mean and volatility respectively, σ_η^2 is the variance of the volatility shock and y_t is a vector of explanatory variables (in our model, y_t is a vector of ones). Both shocks are white noise errors, which are assumed to be distributed independently of each other. The parameter γ measures the sensitivity of the mean variable with respect to the underlying state and is constrained to be positive. The transition probability parameter p_{ij}, where *i* and *j* = {1, 2}, represents the transition probability of going from state *i* to *j*. Note that μ is a function of the latent state s_t, which follows a *k*-state ergodic discrete first-order Markov process as in [11]. The underlying state s_t can assume *k* possible states, that is, one of {1, 2, . . . , *k*}, where higher values of s_t lead to higher intercept terms in the log variance equation. As an identification condition, we require each regime to correspond to at least one time point. A *k*-state stationary transition probability matrix governs the dynamics of the transition from one state to the next state [12]. The latent volatility can be seen as a mixture of *k* densities, where each density corresponds to a single state. The latent volatility at a given time comes from a single density, which is decided by an underlying *k*-state Markov process. This latent regime is denoted by $S_t \in \{1,\ 2,\ 3\}$, which corresponds to the negative jump, regular and positive jump regimes, respectively. Our stochastic volatility model is

$$y_t \middle| \Theta, y_{t-1}, \cdots, y_{t-p}, \sigma_t,\ S_t \sim \mathrm{N}(\mathrm{X}_t' \beta_{S_t}, \sigma_t^2)$$
$$\ln\sigma_t \middle| \Theta, \sigma_{t-1}, \cdots, \sigma_{t-q},\ S_t \sim \mathrm{N}(Z_t' \gamma_{S_t}, \tau_{S_t}^{-1}) \tag{3}$$

The regressors in the mean equation are $\mathrm{X}_t' = (1, y_{t-1}, \cdots, y_{t-p})$ and in the volatility equation $Z_t' = (1, \ln\sigma_{t-1}, \cdots, \ln\sigma_{t-q})$. For notational simplicity, we introduce $\mathrm{T}\times 1$ vectors of $y,\ \sigma$ and $\ln\sigma$, the $\mathrm{T}\times\mathrm{N}$ matrix X and $\mathrm{T}\times\mathrm{M}$ matrix Z. It will also be convenient to collect all T_r observations that belong to regime *r* in the separate vectors and matrices, for *r* = 1, 2, 3. The $\mathrm{T}_r \times 1$ vectors y_r, σ_r and $\ln\sigma_r$, the $\mathrm{T}_r \times \mathrm{N}$ matrices X_r and the $\mathrm{T}_r \times \mathrm{M}$ matrices Z_r contain only those rows of the original vectors and matrices with the $S_t = r$. Finally let $\Sigma = diag(\sigma_t^2)$ be a $\mathrm{T}\times\mathrm{T}$ matrix, and create diagonal $\mathrm{T}_r \times \mathrm{T}_r$ matrices Σ_r similarly. We may then write our stochastic volatility model as

$$y_r \middle| \Theta,\ \sigma_r, S \sim \mathrm{N}\left(\mathrm{X}_r \beta_r, \Sigma_r\right) \quad \ln\sigma_r \middle| \Theta,\ S \sim \mathrm{N}\left(Z_r \gamma_r, \tau_r^{-1} \mathrm{I}_{\mathrm{T}_r}\right) \ \textit{for } r = 1,\ 2,\ 3. \tag{4}$$

We follow [13] and set $\ln\sigma_t \middle| S_t \sim \mathrm{N}(0, \tau_{S_t}^{-1})$ independent for $t = 1-q, \cdots, 0.$ The state dynamics of the transition probabilities $P\left(S_{t+1} \middle| S_t = r, \Theta\right)$ are given implicitly by

$$S_{t+1} = \begin{cases} 1 & \textit{if } \omega_t' \delta_r + \varepsilon_t < 0 \\ 2 & \textit{if } 0 \le \omega_t' \delta_r + \varepsilon_t \le \alpha_r \\ 3 & \textit{if } \alpha_r < \omega_t' \delta_r + \varepsilon_t \end{cases} \tag{5}$$

where ε_t is independent identically distributed N(0, 1), $\omega_t' = \Theta$ and the parameters δ_r and α_r may again be different for each regime.

Following the existing switching literature, we limit ourselves to two states: a high volatility state and a low volatility state, i.e., we set *k* = 2. The RSV model specification combines a level effect and a conditional volatility process that is driven by two shocks, s_t and η_t. The estimation of the RSV model involves estimation of mean parameters $\{a_0,\ a_1\}$, variance parameters $\{\alpha,\ \beta,\ \gamma,\ \sigma_\eta^2,\ \varphi_1\}$, and transition probability parameters $\{p_{01},\ p_{10}\}$. The RSV model reduces to the univariate or Single-state Stochastic Volatility (SSV) model when μ is state independent, that is, when γ is equal to zero. The SSV model also reduces to the [14] model when the conditional volatility is specified as a GARCH process.

For the two-regime (conditionally) independent RSV model, we have six parameters to estimate, $\Theta=\{\beta,\ \gamma,\ \sigma_{\eta}^{2},\ \phi_1,\ p_{01},\ p_{10}\}$ along with the two latent variables, $\Sigma_t^2=\{\sigma_1^2,\ \cdots,\ \sigma_t^2\}$ and $S_t=\{s_1,\ \cdots,\ s_t\}$. The parameter set therefore consists of $\omega=\{\Sigma_t^2,\ S_t,\ \Theta\}$ for all *t*. Bayes theorem is used to decompose the joint posterior density as follows:

$$f\left(\Sigma_n^2,\ S_n,\Theta\right)\alpha\ f\left(\mathrm{Y}_n\middle|\Sigma_n^2\right)f\left(\Sigma_t^2\middle|S_n,\Theta\right)f\left(S_n\middle|\Theta\right)f(\Theta) \tag{6}$$

We draw the marginals $f\left(\Sigma_t^2\middle|\mathrm{Y}_t,S_t,\Theta\right)$, $f\left(S_t\middle|\mathrm{Y}_t,\Sigma_t^2,\Theta\right)$ and $f\left(\Theta\middle|\mathrm{Y}_t,\Sigma_t^2,S_t\right)$ using the Gibbs sampling algorithm. We first draw the underlying volatility $f\left(\Sigma_t^2\middle|\mathrm{Y}_t,S_t,\Theta\right)$ using the multi-move simulation sampler based on [13]. Representing the conditional mean as a mixture of normal variates as in [14], we then draw from the seven underlying normals. We next draw the underlying Markov-state $f\left(S_t\middle|\mathrm{Y}_t,\Sigma_t^2,\Theta\right)$ as in [15]. Then, we cycle through the conditionals of parameter vector Θ for the volatility equation following [16]. For the Gibbs estimation, we set the burn-in iterations as 4,000. We sample from the next 6,000 draws and choose every fifth observation to minimize possible correlation in the draws following [12]. We construct 95% confidence intervals and the standard errors for the parameters. We estimate the density functions for the parameters based on the Gaussian kernel estimator [17].

3 Methodology

The estimation of the RSV model involves estimating two latent variables, i.e., σ_t^2 and s_t, in addition to the model parameters. In the presence of two latent variables, the likelihood function for the model needs to be integrated over all the possible states of the two latent variables. Jacquier, Polson and Rossi [18] show that maximum likelihood-based methods tend to fail under complex specifications of the likelihood function. Consequently, we resort to Monte Carlo Markov Chain (MCMC) methods [19] to estimate the RSV model and Maximum Likelihood Estimation (MLE) for GARCH family of models. The likelihood function is given as: let $\mathrm{Y}_t=\log\frac{S_{t+1}}{S_t}$ be the log returns in (t+1)th day. The likelihood for observation $y=(y_1,y_2,\cdots,y_n)$ is

$$L(\Theta)=f\left((y_1|\Theta)f(y_2|\Theta,y_1)f(y_3|\Theta,y_1,y_2)\cdots f(y_n|\Theta,y_1,\cdots,y_{n-1})\right) \tag{7}$$

where f is the probability density function (pdf) of *y*. Hence, the contribution to the log-likelihood of the tth observation is $\log f\left(y_t\middle|y_{t-1},y_{t-2},\cdots,y_1,\Theta\right)$.

Adapting [20], we calculate recursively for each *t* as

$$f\left(p_t,p_{t-1},y_t\middle|y_{t-1},y_{t-2},\cdots,y_1,\Theta\right)=p\left(p_{t-1}\middle|y_{t-1},\cdots,y_1,\Theta\right)\times p\left(p_t\middle|p_{t-1},\Theta\right)f\left(y_t\middle|p_t,\Theta\right) \tag{8}$$

where $p(p_t|p_{t-1},\Theta)$ is the transition probability between the regimes; $f(y_t|p_t,\Theta)=\phi(y_t-\mu_{s_t})/\sigma_{s_t}$, where ϕ is the standard normal probability density function and the probability $p(p_{t-1}|y_{t-1},\cdots,y_1,\Theta)$ is found from the recursion to be equal to

$$\frac{f\left(p_{t-1},p_{t-2}=1,y_{t-1}\middle|y_{t-2},y_{t-3},\cdots,y_1,\Theta\right)+f\left(p_{t-1},p_{t-2}=2,y_{t-1}\middle|y_{t-2},y_{t-3},\cdots,y_1,\Theta\right)}{f\left(y_{t-1}\middle|y_{t-2},y_{t-3},\cdots,y_1,\Theta\right)} \tag{9}$$

To calculate $f\left(y_t\middle|y_{t-1},y_{t-2},\cdots,y_1,\Theta\right)$ as the sum over the four possible values of the equation (2), that is, for $p_t=0,\ 1$ and $p_{t-1}=0,\ 1$, we start the recursion requiring a value (given Θ) for $p(p_0)$, which we can find from the invariant distribution of the regime-switching Markov chain. The invariant distribution $\pi=(\pi_1,\pi_2)$ is unconditional probability distribution for the process. Under the invariant distribution π, each transition returns the same distribution; that is, $\pi\mathrm{P}=\pi$, given $\pi_1p_{0,1}+\pi_2=0,\ 1;\ \pi_1p_{0,0}+\pi_2p_{1,0}=\pi_1$ and $\pi_1p_{0,1}+\pi_2p_{11}=\pi_2$. Clearly $p_{0,0}+p_{1,0}=1$ so that $\pi_1=p_{1,0}|(p_{0,1}+p_{1,0})$ and similarly, $\pi_2=1-\pi_1=p_{0,1}|(p_{0,1}+p_{1,0})$. Hence, we can start the recursion by calculating for a given parameter set Θ:

$$\begin{aligned} f\left(p_1=0,y_1\middle|\Theta\right)&=\pi_1\phi\left(\frac{y_1-\mu_1}{\sigma_1}\right)\\ f\left(p_1=1,y_1\middle|\Theta\right)&=\pi_2\phi\left(\frac{y_1-\mu_2}{\sigma_2}\right)\\ f\left(y_1\middle|\Theta\right)&=f\left(p_1=0,y_1\middle|\Theta\right)+f\left(p_1=1,y_2\middle|\Theta\right) \end{aligned} \tag{10}$$

and we calculate for use in the next recursion the two values of

$$f\left(p_1\middle|y_1,\Theta\right)=\frac{f\left(p_1,y_1\middle|\Theta\right)}{f\left(y_1\middle|\Theta\right)}. \tag{11}$$

Maximizing the likelihood function over the six parameters may be done with standard search methods.

4 Results and Materials

4.1 Empirical Analysis of data

The data consist of weekly observations of First Bank of Nigeria (FBN) stock market price data for the period 01/06/07 to 03/04/18 (528 weekly observations excluding public holidays) are obtained from the Nigeria Stock Exchange's database.

Table 1 presents the summary statistics of the data. Changes in yields, Δr_t, seem to be left-skewed, indicating that yield increases were less common than yield decreases on a weekly basis. There is also strong evidence of kurtosis in the return series. The kurtosis is greater than three for the stock markets and we have leptokurtic distribution. Importantly, the Jarque-Bera statistic is too large and the probability is zero for the stock market price data. Therefore, we reject the null hypothesis that the series follow a normal distribution. The Ljung-Box statistic suggests that there is a high degree of autocorrelation for the raw yields (r_t). On the other hand, Δr_t series seems to be much less persistent and is characterized by low autocorrelations. Looking at the Ljung-Box statistic for the squared residuals (RES_t^2) at various lags, the null of no ARCH effects is strongly rejected by the data. This indicates high autocorrelations in the data that imply time dependence in higher order moments.

Table 1: Descriptive Statistics

	r_t	Δr_t	$(\Delta r_t)^2$	$\log(\Delta r_t)^2$
Mean	5.018(0.58)	0.000(0.007)	0.061(0.003)	-5.832(0.054)
Variance	8.337(0.326)	0.048(0.002)	0.035(0.009)	7.184(0.202)
Skewness	1.391(2.068)	-1.091(0.005)	9.317(0.023)	-0.317(0.695)
Kurtosis	5.732(19.826)	18.156(0.008)	79.653(0.051)	2.891 (6.183)
Ljung-Box(24)	1854.1	13.027	101.28	193.37
LB-ARCH(24)	-	113.58	-	-
Jarque-Bera	1889.2	5876.4	1837.5	1819.6
Probability	0.0000	0.0000	0.0000	0.0000

LB: Ljung-Box statistics is calculated with 24 lags. The χ^2_{24} critical value is calculated for a 95% confidence level of 36.4151. LB-ARCH: Ljung-Box statistic is reported for the square residuals at lag 24, where residuals are obtained from regressing Δr_t on a constant and r_{t-1}.

The stationarity of the variable of both the natural logarithms of the series and the returns were tested using the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. The unit root test results are presented in Table 2.

Table 2: unit Root Tests of the Series and Returns

	Logs	Returns
ADF t-statistics	-1546	-9.073***
KPSS LM-stat	2.358***	0.067

*** denotes rejection of null hypothesis at 1% significance level.

The null hypothesis for the ADF test is that the variable we examine has a unit root since the t-statistic is greater than the critical values for the variable, we cannot reject the null hypothesis. Thus, the variable has a unit root and it is not stationary. In the returns, we reject the null hypothesis at 1% level of significance as the t-statistics is smaller than the critical values for the variable and we have stationarity. The null hypothesis for the KPSS test differs from that of the ADF test and the null hypothesis now is that the variable is stationary. The Lagrange Multiplier (LM) stat for the logarithms is greater than the critical values and we reject the null hypothesis at 1% significance level, so the variable has a unit root. In the returns, the LM-stat is smaller than the critical values and we have stationarity. So both tests reach the same inference. All our series are integrated of order one (I(1)).

For the purpose of estimation and comparison to alternative volatility models, the mean adjusted version of the RSV model is given as:

$$
\begin{aligned}
&\Delta r_t - (\hat{a}_0 + \hat{a}_1 r_{t-1}) \equiv RES_t \\
&RES_t = \sigma_t r_{t-1}^{2\alpha} \varepsilon_t, \qquad \alpha = 0.5 \\
&(\ln(\sigma_t^2) - \mu_{s_t}) = \phi_1(\ln(\sigma_{t-1}^2) - \mu_{s_{t-1}}) + \sqrt{\sigma_\eta^2}\,\eta_{t-1} \\
&\mu_{s_t} = \beta + \gamma s_t \quad \gamma > 0 \qquad s_t = \{1,\ 2\} \\
&\mathrm{P}[s_t = j | s_{t-1} = i] = p_{ij}
\end{aligned} \tag{12}
$$

where all the assumptions on the error terms made in (12) still hold. To benchmark our results, first, we ignore the possibility of regime-switching in the data. When we set γ to zero, the RSV model (12) reduces to the SSV model.

$$
\begin{aligned}
&\Delta r_t - (\hat{a}_0 + \hat{a}_1 r_{t-1}) \equiv RES_t \\
&RES_t = \sigma_t r_{t-1}^{2\alpha} \varepsilon_t, \qquad \alpha = 0.5 \\
&(\ln(\sigma_t^2) - \mu) = \phi_1(\ln(\sigma_{t-1}^2) - \mu) + \sqrt{\sigma_\eta^2}\,\eta_{t-1} \\
&\mu_{s_t} = \beta
\end{aligned} \tag{13}
$$

The results from the MCMC estimation of the SSV model are presented in Table 3, where the parameter set is $\Theta = \{\beta,\ \phi,\ \sigma_\eta^2\}$.

Table 3: Estimation of SSV Model

Parameters	Prior Values		Posterior Values		
	Mean	Std. Dev	Mean(Std. error)	Std. Dev	95% Confidence Interval
β	0.07	1	2. 925(0.031)	0.348	(2.478, 3.763)
ϕ	0	11	0.874(0.000)	0.008	(0.875, 0.976)
σ^2	-	-	0.278(0.015)	0.025	(0.263, 0.348)

The standard errors are reported in parenthesis.

The persistence parameter ϕ is very high, indicating that the half-life of a volatility shock, measured as $-\frac{\ln(2)}{\ln(\phi)}$, is about 15 weeks. Standard errors for the parameters are small, indicating that parameters are highly significant. Figure 1 plots the posterior densities of the parameters. All the parameters have symmetric densities, while half-life density is right-skewed (with mean and median of, respectively, 14.92 and 13.58). That is, we are more likely to observe half-lives less than the mean value of 15 weeks.

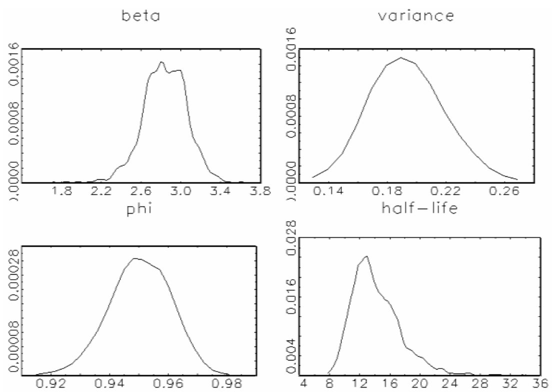

Fig. 1: Posterior Density Plots for Parameters of the SSV Model

Table 4 presents the prior and posterior parameter estimates of the parameter set Θ in our model, where $\Theta = \{\beta,\ \gamma,\ \phi,\ \sigma_\eta^2,\ p_{01}, p_{10}\}$.

Table 4: Estimation of RSV Model

Parameters	Prior Values		Posterior Values		
	Mean	Std. Dev	Mean(Std. error)	Std. Dev	95% Confidence Interval
β	0.07	49	2. 592(0.001)	0.096	(2.267, 2.816)
γ	1	49	2.687(0.024)	0.258	(2.245, 3.195)
ϕ	0	1	0.569(0.001)	0.050	(0.618, 0.729)
σ^2	-	-	0.872(0.015)	0.135	(0.818, 1.328)
p_{01}	0.25	0.17	0.007(0.001)	0.005	(003, 0.021)
p_{10}	0.25	0.17	0.041(0.001)	0.022	(0.016, 0.074)

Table 4 shows that high volatility states tend to be associated with higher long-run mean of $\ln(\sigma_t^2)$ compared to low volatility states. Standard errors for the parameters are small, as before. The persistence parameter, ϕ, drops significantly to 0.569 from 0.874 in the SSV model. This implies that a switch in the latent regime creates a high persistence in volatility and confirms the earlier results in the literature. The distribution of ϕ is left-skewed with mean and median of 0.569 and 0.588 respectively implying that persistence greater than 0.569 is more common. The transition probabilities, p_{00} and p_{11}, are estimated as 0.993 and 0.959. These estimates are comparable to 0.9896 and 0.9739, respectively, reported in [21]. The analyses imply that the effect of a volatility shock is much more persistent in the low volatility state than in the high volatility state. A volatility shock lasts about 59 weeks in the low volatility state compared to about 18 weeks in the high volatility state, where duration of the shock in state *i* is obtained as $(1-p_{ii})^{-1}$.

Figure 2 plots the densities for the posterior parameter estimates using a Gaussian kernel. The posterior densities seem to be right-skewed for σ_η^2, p_{01} and p_{10} (with medians 0.751, 0.004 and 0.0299, respectively) and symmetric for β and γ. The correlations between the parameters of SSV and RSV models have a strong negative correlation (-0.638 and –0.851 respectively) between ϕ and

σ_η^2, suggesting that there is a tradeoff between volatility persistence and volatility of volatility. The RSV model is also characterized by strong positive correlations between β and γ (0.611) and β and σ_η^2 (0.507) and a negative correlation between β and ϕ (-0.47). These findings suggest that high volatility periods are associated with high volatility regimes, high kurtosis and low persistence. The finding of low persistence in high volatility periods is consistent with [21].

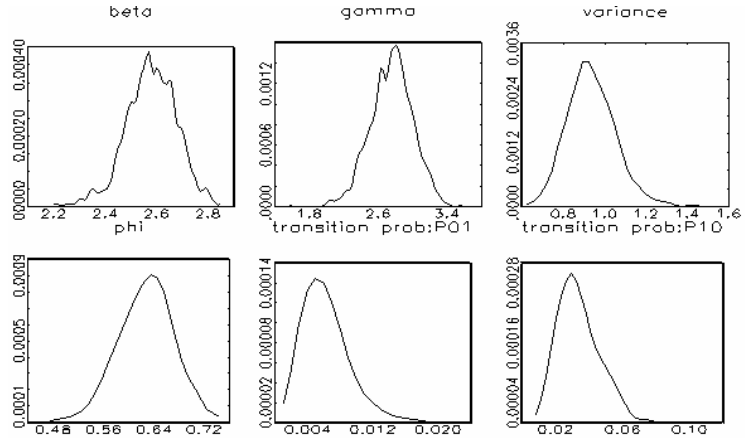

Fig. 2: Posterior Density Plots for Parameters of the RSV Model

4.2 In-sample forecasts

The in-sample performance of the SV models in comparison to the GARCH family of models is considered. The Markov switching ARCH model, proposed by [20], with two states and one autocorrelation lag or SWARCH(2,1) model; the three popular GARCH models with normally distributed innovations: a GARCH(1,1) model [22]; a GARCH(1,1)-L model [22] (i.e., GARCH(1,1) with an asymmetry effect of negative lagged error, to capture the leverage effect) and an EGARCH(1,1) model [23] are considered. All the GARCH models are specified to include a level effect as given:

(i) GARCH (1, 1): $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \qquad t>1$; $\sigma_1^2 = \dfrac{\alpha_0}{1-\alpha_1-\beta_1} \qquad t=1$.

(ii) EGARCH (1, 1): $\ln(\sigma_t^2) = \alpha_1 + \delta_1\left(\left|\varepsilon_{t-1}^2\right| - \sqrt{\dfrac{2}{\pi}}\right) + \delta_2\varepsilon_{t-1} + \beta_1(\sigma_{t-1}^2) \qquad t>1$; $\ln(\sigma_1^2) = \dfrac{\alpha_0}{1-\beta_1} \qquad t=1$.

(iii) GARCH (1, 1)-L: $\sigma_t^2 = \alpha_0 + kd_{t-1}\varepsilon_{t-1}^2 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \qquad t>1$; $\sigma_1^2 = \dfrac{\alpha_0}{1-\dfrac{k}{2}-\alpha_1-\beta_1} \qquad t=1$.

$$d_{t-1} = \begin{cases} 0 & if\ \varepsilon_{t-1} > 0 \\ 1 & if\ \varepsilon_{t-1} \le 0 \end{cases}$$

(iv) SWARCH (1, 1): $\sigma_t^2 \big/ \gamma(s_t) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 \big/ \gamma s_{t-1} \qquad t>1$ where $s_t = 1,\ 2$.

$\gamma(s_t)$ is the multiplicative factor in each state. $\gamma(s_t=1)=1; \quad \mathrm{P}\left[s_t = j \middle| s_{t-1} = i\right] = p_{ij}$.

The maximum likelihood estimation results for the three GARCH models are reported in Table 5.

Table 5: Estimation of ARCH Models

Parameter / Model	α_0	k	γ δ_1	δ_2	α_1	β_1	
GARCH (1, 1)	0.795(-5.13)	-	-	-	-	0.021(-7.233)	0.862(-47.001)
GARCH (1, 1)-L	0.801(-5.22)	0.010(-2.587)	-	-	-	0. 017(-6.437)	0.858(-46.891)
EGARCH (1, 1)	0.052(2.714)	-	-	0.153(-11.621)	-0.018(-3.425)	-	0.923(-127.75)
SWARCH (2, 1)	13.146(14.295)	-	10.285(-10.537)	-	-	-	0.029(-5.138)

t-statistics are reported in parenthesis. The general model used is $\Delta r_t - (\hat{a}_0 + \hat{a}_1 r_{t-1}) \equiv RES_t$

$RES_t = \sigma_t r_{t-1}^{2\alpha}\varepsilon_t, \quad \alpha = 0.5 \qquad (\varepsilon_t|\Omega_{t-1}) \sim \mathrm{N}(0,\ 1)$

There is evidence for a leverage effect based on the significant t-statistic for *k* in the GARCH (1,1)-L model and the significant t-statistic for δ_2 in the EGARCH(1,1) model. The leverage effect, however, is small relative to the usual size found in equity returns.

All the estimates in the conditional variance equation are significant for the three models indicating high persistence in the conditional variance.

The predictive accuracy of the models is evaluated in Table 6 using the model diagnostics.

Table 6: In-Sample Performance

Model	Parameters	Log-likelihood	AIC	SBC	Adj. R^2	Posterior Odds ratios	MSE	JB
Const. Variance	1	-8678.13	-8679.13	-8682.02	-0.479	-	1.3	17467.03*
GARCH(1, 1)	3	-6854.72	-6859.79	-6868.17	0.224	1813.85	0.89	647.91*
GARCH(1, 1)-L	4	-6853.09	-6857.04	-6868.20	0.207	1813.82	1.06	625.87*
EGARCH(1,1)	4	-6842.34	-6846.13	-6859.46	0.222	1822.56	0.45	933.01*
SWARCH(2, 1)	5	-6851.97	-6854.69	-6870.88	0.224	1811.14	1.05	18.20*
SSV Model	3	-6779.38	-6782.75	-6786.45	0.450	1895.57	0.33	231.69*
RSV Model	6	-6281.15	-6285.15	-6296.37	0.611	2385.65	0.28	4.59

Const. Variance stands for the constant variance model with level effect given as: $\Delta r_t - (\hat{a}_0 + \hat{a}_1 r_{t-1}) \equiv RES_t$

$RES_t = \sigma_t r_{t-1}^{2\alpha}\varepsilon_t, \quad \alpha = 0.5 \qquad (\varepsilon_t|\Omega_{t-1}) \sim \mathrm{N}(0,\ 1)$. $Adj.R^2$: Calculated for the regression $RES_t^2 = a + b\sigma_t^2 + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,\ 1)$ and $\{t = 1, \ldots, T\}$ where RES_t are the OLS residuals. σ_t^2 refers to conditional volatility at time *t*. Posterior Odds Ratio: The difference of the SBC of each model and the SBC of the constant Variance model. MSE: Mean squared error defined as $\mathrm{T}^{-1}\sum_{t=1}^{\mathrm{T}}[RES_{t+1}^2 - \sigma_{t+1}^2]^2$. JB: Jarque Bera's normality test statistic, where * indicates significance at 5% level. AIC is Akaike Information Criterion and SBC is Schwartz Bayesian Criterion.

A constant variance model was included as a benchmark for conditional volatility models. Importantly, the SV models have higher log-likelihood values, adjusted R^2s and AIC/SBC values compared to the competing ARCH models. Within the SV models, RSV fares better than SSV based on these metrics. The posterior odds ratio captures the relative performance of each model with respect to the constant variance model [24-25]. If the odds ratio for a model is positive, then that model is "more likely" to have generated the data than the constant variance model. The model with the highest value of posterior odds ratio represents the "most likely" model specification. The stochastic volatility models, in general, have higher odds ratios than the ARCH models. Within SV models, the RSV performs better. In particular, the RSV model has an odds ratio that is at least 75% higher than the other competing models.

Table 6 also presents the mean squared error (MSE) and the Jarque-Bera normality test statistic (JB) for alternative models. The RSV model has the lowest mean squared error (MSE) values, closely followed by the SSV model. The JB statistic rejects the normality assumption for all the models except the RSV model. Among the GARCH models, with the exception of the adjusted R^2 criteria, the EGARCH(1,1) performs better than the other models for all the evaluation measures. The EGARCH(1,1) has an overall better performance than the SWARCH(2,1) model. Based on these considerations, we select the EGARCH(1,1) model for subsequent out-of-sample tests of the SV models.

The differences between conditional volatilities based on their statistical significance were done following [26] who propose two non parametric tests to test the null hypothesis of the equality of forecasts from two competing models: the sign test and the Wilcoxon signed rank test. The signed test is calculated as follows:

$$S_{1a} = \frac{S_1 - .5\mathrm{T}}{\sqrt{.25\mathrm{T}}} \overset{a}{\sim} \mathrm{N}(0,1) \tag{14}$$

where $S_1 = \sum_{t=1}^{\mathrm{T}} \mathrm{I}(d_t)$ and $\mathrm{I}(d_t) = \begin{cases} 1 & if\ d_t > 0 \\ 0 & Otherwise \end{cases}$

where d_t is the loss differential, defined as the difference between the forecast errors generated by the two competing models. The Wilcoxon signed rank test is calculated as follows:

$$S_{2a} = \frac{S_2 - .52\mathrm{T}(\mathrm{T}+1)}{\sqrt{\frac{.\mathrm{T}(\mathrm{T}+1)(2\mathrm{T}+1)}{24}}} \overset{a}{\sim} \mathrm{N}(0,1) \tag{15}$$

where $S_2 = \sum_{t=1}^{\mathrm{T}} \mathrm{I}(d_t) Rank(|d_t|)$ and $\mathrm{I}(d_t) = \begin{cases} 1 & if\ d_t > 0 \\ 0 & Otherwise \end{cases}$

Table 7: RSV versus Other Competing Models

	Sign Test	Wilcoxon Signed Rank Test
Const. Variance	-25.163*	-17.727*
GARCH (1, 1)	-17.029*	-14.999*
GARCH (1, 1)-L	-16.917*	-14.543*
EGARCH (1,1)	-17.118*	-13.918
SWARCH (1, 1)	41.659*	36.433*
SSV Model	3.089	4.013

The test statistic for Sign test and Wilcoxon Signed Rank test follows a standard normal distribution.
* indicates significance at 5% level.

Table 7 presents results for the standard normal z statistic for the two tests. The RSV model forecasts are significantly different from the forecasts of the competing models for the in-sample period at 5% level. The negative sign of the test statistic implies that the in-sample volatility forecasts from the RSV model are much lower than those from the constant volatility and other GARCH models. The SSV model has lower in-sample volatility forecasts compared to the RSV model. The in-sample results are very supportive of the RSV model probably due to the richer parameterization of the RSV model relative to the other models; over-fitting might play a part in the in-sample success of the RSV model. The in-sample results show that the SV models are superior to the ARCH type models based on different metrics. The SV models have significantly lower conditional volatility estimates compared to the ARCH models. Within the ARCH models, the EGARCH(1,1) has the best in-sample performance. Thus, to better judge our RSV model, we perform an out-of-sample evaluation of all the models.

4.3 Out-of-sample forecasts

Table 8 reports the out-of-sample performance of the four models. The constant volatility model, the best performing ARCH model based on the in-sample period, i.e., the EGARCH(1,1) and both SV models are considered.

Table 8: In-sample and out-of-sample volatility comparison of alternative models

	In-sample (T = 282) 6/1/07 – 31/12/11		Out-of-Sample (fixed sample) 1/1/12 -3/4/18	
	MSE	MAE	MSE	MAE
Sample 1 Const. Variance	0.084	0.0572	0.0003	0.0069
EGARCH (1,1)	0.0453	0.0519	0.0002*	0.0051
SSV Model	0.0456	0.0519	0.0002*	0.0049*
RSV Model	0.0450*	0.0513*	0.0002*	0.0049*
	In-sample (T = 220) 1/1/12 – 31/12/16		Out-of-Sample (fixed sample) 1/1/17 – 3/4/18	
	MSE	MAE	MSE	MAE
Sample 2 Const. Variance	0.1214	0.1162	0.0004*	0.0128
EGARCH (1,1)	0.1185	0.1153	0.0007	0.0137
SSV Model	0.1173	0.1141	0.0007	0.0125
RSV Model	0.1122*	0.1105*	0.0004*	0.0116*

* indicates best model. MAE is the mean absolute error defined as $\mathrm{T}^{-1}\sum_{t=1}^{\mathrm{T}}\left|RES_{t+1}^{2}-\sigma_{t+1}^{2}\right|$.

The estimated coefficients from the in-sample period are used to generate one-week (step) ahead conditional volatility estimates for the out-of-sample period. Using estimates for the in-sample period 1 to *t*, we generate one-step-ahead conditional volatility forecasts for each future time period t + k, where $\{k = 1,\ 2,\ 3,\ \cdots,\ \mathrm{T}-t,\ \mathrm{T}$ is the sample size$\}$. The RSV model tends to perform better than the EGARCH (1,1) (except in sub-sample 1, where the MSEs are the same for both models). In fact, the EGARCH model never performs better than the SV models out-of-sample in terms of MAE. The out-of-sample performance of the RSV model and the SSV model in sub-sample 1 is both similar. The results are not surprising because the less switching in the out-of-sample period, the less efficient the RSV model should be relative to the SSV model as indicated in sub-sample 1 with no high volatility regimes. That is, the RSV model performs better than the SSV model whenever the out-of-sample period has regimes switching between low and high volatility. Consistent with [27, 12], the constant variance model shows a good out-of-sample performance, especially in the MSE metric. Note that the constant variance model in the first sub-sample beats all the other models. The second sub-sample presents out-of-sample forecasts for approximately one year period. Again, the RSV model shows superior performance. The out-of-sample results show that the SV models generally outperform the best performing ARCH model, the EGARCH (1,1) model.

5 Conclusion

In this paper, we model the volatility of market stock prices as a stochastic volatility process whose mean is subject to shifts in regime. The analysis shows that Markov switching models provide a flexible framework to handle many features of asset returns. In particular, they allow for nonlinearities arising from persistent jumps in the model parameters. These models have several appealing features. First, they provide a convenient framework to endogenously identify regime shifts that are common place in financial data;

the usual high volatility persistence is substantially reduced by the introduction of regime-switching. Second, as Markov switching models belong to the mixture-of-distributions class of stochastic processes, they are as versatile as mixture models in capturing salient features of financial data such as mean reversion, time-varying volatilities, skewness, and leptokurtosis. The results also indicate that stochastic volatility models (SSV and RSV) outperform the ARCH models both in-sample and out-of-sample. Our results are consistent with those of [28] and [29], who find that the stochastic volatility models typically generate lower historical volatilities and hence lower option prices than constant volatility and GARCH models.

References

[1] Musiela, M. and Rutkowski, N. (2005). Martingale Methods in Financial Modelling, Springer.

[2] Karakatsani, N. V. and Bunn, D. W. (2008). Intraa-day and regime-switching dynamics in electricity price formation. *Energy Economics*, 30: 1776-1797.

[3] Onyeka-Ubaka, J. N. (2013). *A BL-GARCH Model for Distribution with Heavy Tails.* A Ph.D. Thesis, Department of Mathematics, University of Lagos, Nigeria.

[4] Durham, G. B. (2007) SV mixture models with application to S&P 500 index returns. *Journal of Financial Economics*, 85(3): 822–856.

[5] Onyeka-Ubaka, J. N., Abass, O. and Okafor, R. O. (2014). Heavy Tails Estimation in Nonlinear Models. *International Journal of Statistics and Applications*. DOI: 10.5923/j.statistics.20140401.02, 4(1):18-27.

[6] Renault E. and Touzi, N. (1996) Option Hedging and Implied Volatilities in a Stochastic Volatility Model. *Mathematical Finance*, 6(3):279–302.

[7] Trolle, A. B. and E. S. Schwartz, E. S. (2009). A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives. *The Review of Financial Studies*, 22 (5).

[8] Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics*, 12: 309-316.

[9] Andreasen, J. and Dahlgren, M. (2006). At the Flick of a Switch. *Energy Risk*, 10: 71-75.

[10] Geman, H. and Roncoroni, A. (2006). Understanding the Fine Structure of Electricity Prices, *Journal of Business*, 79: 1225-1261.

[11] Hamilton, J. D. (1988). Rational expectations econometric analysis of changes in regime. *Journal of Economic Dynamics and Control*, 12: 385-423.

[12] Nitsakou, A. (2016). *The predictive power of regime-switching models for stock market returns* Interdepartmental Programme of Postgraduate Studies in Economics (Masters in Economics) Thesis, Thessaloniki.

[13] De Jong, P. and Shephard, N. (1995). Simulation smoother for time series models. *Biometrika* 82: 339-50.

[14] Brenner, R. J., Harjes, R. and Kroner, K. (1996). Another look at models of short-term interest rates. *Journal of Financial and Quantitative Analysis* 31: 85-107.

[15] Carter, C. K. and Kohn R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81: 541-53.

[16] Albert, J. H. and Chib, S. (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to regime shifts. *Journal of Business Statistics and Economics*, 11: 1-15.

[17] Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman and Hall, New York.

[18] Jacquier, E., Polson, N. G. and Rossi, P. E. (1995). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12: 371-392.

[19] Eraker, B. (2001). MCMC analysis of diffusion models with applications to finance. *Journal of Business and Economic Statistics*, 19(2): 177-191.

[20] Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroscedasticity and changes in Regime. *Journal of Econometrics*, 64: 307-333.

[21] Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime switching process. *Journal of Financial Economics*, 42: 27-62.

[22] Bollerslev, T. (1986).Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31: 307-328.

[23] Nelson, D.B. (1991). Conditional heteroskedasticity in Asset returns: A new Approach. *Econometrica*, 59: 347- 370.

[24] Kim. D. and Kon, S. J. (1994). Alternative models for the conditional heteroscedasticity of stock returns. *Journal of Business*, 67: 563-98.

[25] Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65: 361-94.

[26] Diebold, F. X. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics,* 13: 253-265.

[27] Figlewski, S. (1997). Forecasting volatility. Financial Markets, Institutions and Instruments 6, 1, New York University Salomon Center.

[28] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2): 357-384.

[29] Vetzal, K. R. (1997). Stochastic volatility, movements in short term interest rates and bond option values. *Journal of Banking and Finance*, 21:169-96.