Forecasting Volatility of Stock Indices with HMM-SV Models

Nkemnole E. B.¹*, Abass O.²**
*Department of Mathematics, University of Lagos, Nigeria. enkemnole@unilag.edu.ng ¹,
**Department of Computer Sciences, University of Lagos, Nigeria. olabass@unilag.edu.ng ²

Abstract

The use of volatility models to generate volatility forecasts has given vent to a lot of literature. However, it is known that volatility persistence, as indicated by the estimated parameter $\phi$, in Stochastic Volatility (SV) model is typically high. Since future values in SV models are based on the estimation of the parameters, this may lead to poor volatility forecasts. Furthermore, this high persistence, as contended by some writers, is due to the structure changes (e.g. shift of volatility levels) in the volatility processes, which SV model cannot capture. This work deals with the problem by bringing in the SV model based on Hidden Markov Models (HMMs), called HMM-SV model. Via hidden states, HMMs allow for periods with different volatility levels characterized by the hidden states. Within each state, SV model is applied to model conditional volatility. Empirical analysis shows that our model, not only takes care of the structure changes (hence giving better volatility forecasts), but also helps to establish an proficient forecasting structure for volatility models.

Keywords: Forecasting, Hidden Markov model, Stochastic volatility, stock exchange

1.0 Introduction

A great deal of attention has been paid to modeling and forecasting the volatility of stock exchange index via stochastic volatility (SV) model in finance, as well as empirical literature. No doubt, forecasting the volatility of stock exchange index is an important aspect of many financial decisions. For instance, investment managers, option traders and the financial managerial bodies are all interested in volatility forecasts in order to either construct less risky portfolios or obtain higher profits (Panait and Slavescu, 2012). Hence, there is always a need for good analysis and forecasting of volatility.
Various volatility models have been put forward to describe the statistic features of financial time series. Lots of documents which explain the sources of the market volatility have been presented by Shiller (1993). In the words of Bob Jarrow, “many of these models come out of the academic community, where disagreements over the latest and greatest models have become something of an armchair sport”. SV model are the trendiest. Other volatility models include the autoregressive conditional heteroskedasticity (ARCH) models by Engle (1982) and extended to generalized ARCH (GARCH) by Bollerslev et al. (1995). Their success lies in their ability to capture some stylized facts of financial time series, such as time-varying volatility and volatility clustering.

Many financial variables, such as stock prices and exchange rate exhibit volatility clustering whereby volatility is likely to be high when it has recently been high and volatility is likely to be low when it has recently been low. GARCH models are particularly valuable for modelling time-varying conditional volatility; they are extensively used by both researchers and practitioners. GARCH models the time varying variance as a deterministic function of lagged squared residuals and lagged conditional variance. Also, volatility is usually predicted by using the GARCH model (Bollerslev et al. 1995).

SV model is another way of modeling time varying volatility (Taylor, 1982, 1986). SV models the variance as an unobserved component that follows a particular stochastic process. In it, time-varying variance is not restricted to a fixed process. In SV models it is usual to model volatility as a logarithmic first order autoregressive process. It therefore represents a discrete time approach to the diffusion process used in the option pricing literature (Hull and White, 1987). This model, though theoretically attractive, is empirically challenging as the unobserved volatility process enters the model in a non-linear fashion which leads to the likelihood function depending upon high-dimensional integrals. A number of varied approaches can be used to estimate the SV model. Two recent studies that compare the usefulness of the SV model with GARCH models in applied forecasting situations can be seen in So et al. (1999) and Yu (2002). So et al. (1999), for instance, ascertained that in modelling and forecasting foreign exchange rates, the SV model estimated as a state space model does not, in general, outperform GARCH model. Yu (2002), on his own part, used the SV model to forecast daily stock market volatility for New Zealand. By means of forecast accuracy tests, he discovered that the SV model
surpasses performance of GARCH models. The mixed results from these two papers suggest the need for further research on the relative merits of SV models in applied forecasting situations. Nelson (1991) and Glosten, et. al., (1993) have used the GARCH model to compute the difference, which has effects of negative and positive shocks on volatility. Kim, et. al. (1998), in recent times, applied Hidden Markov model (HMM), instead of ARCH, to handle the effects in economic data. The difference between the HMM and ARCH is the unconditional variance. If there are sequential changes in regime, some researchers advise that some more intuitive approaches need to be considered, and using different regimes may contribute to the return-generating process in the market. Hamilton and Susmel (1994) apprehended that the long run variance could obey regime shift; they suggested an ARCH process. The effect will vanish if they use weekly data, because sparse time point makes the dependence weaker. It is rational to examine the price of stock market by using HMM. In using HMM, Chu, et. al., (1996) chose a two-stage process to represent the return behavior in the stock market. They first considered the return behavior in stock market as a Markov process. Then, the different return regimes derived from the first stage were utilized to estimate the volatility. Lastly, they found that the negative deviations in returns can have larger increase in volatility than the positive one. Accordingly, they think the return and volatility are not linearly but asymmetrically. HMM have been applied for at least three decades in signal-processing applications, especially in automatic speech recognition. Now this theory and application has expanded to other fields. A HMM (see Rabiner and Fellow (1989)) includes two stochastic processes of which one is an underlying stochastic process that is not observable and, the other process is the observation sequence.

Although standard SV models improve the in-sample fit a lot compared with constant variance models, numerous studies find that SV models give unsatisfactory forecasting performances, (Figlewski, 1997). Xiong-Fei , and Lai-Wan (2004) argued that the usually overstated volatility persistence in SV models may be the cause of poor forecasting performances. For Lamoureux (1990), this well-known high persistence may originate from the structure changes in the volatility processes, which SV models cannot capture. He demonstrated that any shift in the structure of financial time series (e.g. the shift of unconditional variance) is likely to lead to misestimating of the SV parameters in such a way that they entail too high a volatility persistence.
In this paper, we intend to solve the problem of excessive persistence in SV model by bringing in HMM to allow for different volatility states (periods with different volatility levels) in time series. And also, within each state, we allow SV model to model the conditional variance. The ensuing HMM-SV model indeed yields better volatility forecast compared to SV models for artificial data and real financial data, in-sample as well as out-of-sample.

2. HMM-SV model

2.1 Hidden Markov Model
Although initially introduced and studied as far back as 1957 and early 1970’s, the recent popularity of statistical methods of HMM is not in question. A HMM is a bivariate discrete-time process \( \{X_k, Y_k\}_{k \geq 0} \) where \( \{X_k\}_{k \geq 0} \) is an homogeneous Markov chain which is not directly observed but can only be observed through \( \{Y_k\}_{k \geq 0} \) that produce the sequence of observation. \( \{Y_k\}_{k \geq 0} \) is a sequence of independent random variables such that the conditional distribution of \( Y_k \) only depends on \( X_k \). The underlying Markov chain \( \{X_k\}_{k \geq 0} \) is called the state.

The model, for MacKay (2003) is used for two purposes namely, (1) to make inferences about an unobserved process based on the observed one and (2) to explain variation in the observed process based on variation in a postulated hidden process. Rabiner (1989) in one of his papers in HMM, portrayed this technical trend under two reasons:

“First the models are very rich in mathematical structure and hence can form the theoretical basis for use in a wide range of applications. Second the models, when applied properly, work very well in practice for several important applications”.

HMM are equivalently defined through a functional representation known as state space model. The state space model (Doucet and Johansen, 2009) of a HMM is represented by the following two equations:

\[
\begin{align*}
(\text{State equation}) \quad x_t &= f(x_{t-1}) + w_t \\
(\text{Observation equation}) \quad y_t &= g(x_t) + v_t
\end{align*}
\]

(1) (2)
where \( f \) and \( g \) are either linear or nonlinear functions, while \( w_i \) and \( v_i \) are white noise processes. Models represented by (1) - (2) are referred to as state space model and this include a class of HMMs with non-linear Gaussian state-space model such as stochastic volatility (SV) model.

A standard hidden Markov model (Zhang et al. 2007) is characterized as follows

\[
X_t \sim f(X_t \mid X_{t-1}, A) \tag{3}
\]

\[
Y_t \sim g(Y_t \mid X_t, B) \tag{4}
\]

where \( X_t \in S \) is a hidden state at time \( t \), \( S = \{S_1, S_2, \ldots\} \).

We now define the notation of an HMM Rabiner (1989) which will be used later. Given the time series \( Y \), an HMM is characterized by the following:

1. \( N \), the number of states in the model. In our model the states refer to different variance levels. We denote the state set as \( S = \{S_1, S_2, \ldots, S_N\} \), and the state at time \( t \) as \( x_t \), \( x_t \in S \)

2. The state transition probability distribution \( A = \{a_{ij}\} \) where

\[
a_{ij} = p(x_t = S_j \mid s_{t-1} = S_i) , \quad a_{ij} \geq 0, 1 \leq i, j \leq N .
\]

3. The observation probability distribution \( B. \) \( Y_t \sim g(Y_t \mid X_t, B) \) is observation density at time \( t \), \( B = \{b_j(k)\} \) gives the conditional probability distribution of each observation symbol within a given hidden state with definition of

\[
b_j(k) = p(o_t = v_k \mid q_t = j) , 1 \leq j \leq N , 1 \leq k \leq M , v_k \text{ denotes the } k^{th} \text{ observation symbol per state.}
\]

4. The initial state distribution \( \pi = \{\pi_i\} , \pi_i = p(x_1 = s_i) , 1 \leq i \leq N \). For convenience, we used the compact notation \( \lambda = (A, B, \pi) \) to indicate the complete parameter set of the model. Given the form of the HMMs, the goal is to find the best model for a given time series through optimally adjusting model parameters \( \lambda = (A, B, \pi) \).

Given the model and the observation sequence, the model parameter is estimated with the following estimation algorithm. The first two are pattern recognition problems: Finding the probability of an observed sequence given a HMM (evaluation); and finding the sequence of
hidden states that most probably generated an observed sequence (decoding). The third problem is generating a HMM given a sequence of observations (learning). It deals with the training of the model which is of most significant interest.

Evaluation: Given a model \( \lambda = (A, B, \pi) \), and a sequence of observations \( O = (o_1, \ldots, o_T) \), how do we compute \( p(O | \lambda) \)? We use the forward algorithm to calculate the probability of an observation sequence given a particular HMM.

The forward variable \( \alpha_t(i) \) is defined as:

\[
\alpha_t(i) = P(o_1 o_2 \cdots o_t, q_t = s_i | \lambda)
\]  

(5)

\( \alpha_t(i) \) stores the total probability of ending up in states \( s_i \) at time \( t \), given the observation sequence \( o_1, o_2, \ldots, o_t \), then the sum of \( \alpha_t(i) \) gives the probability of the observation, given the HMM, \( \lambda \).

\[
P(O | \lambda) = \sum_{i=1}^{N} \alpha_T(i)
\]  

(6)

The forward variable at each time \( t \) is calculated inductively as follows:

1. Initialisation \( \alpha_1(i) = \pi_i b_i(o_1), \ 1 \leq i \leq N \)

2. Induction \( \alpha_{t+1}(j) = \left[ \sum_{i=1}^{N} \alpha_t(i)a_{ij} \right] b_j(o_{t+1}), 1 \leq t \leq T - 1, 1 \leq j \leq N \)

3. Update time set \( t = t + 1 \); Return to step 2 if \( t < T \); else terminate algorithm.

4. Termination \( P(O | \lambda) = \sum_{i=1}^{N} \alpha_T(i) \)

Full details of the procedure as well as the various implementation issues, are described in Bhar and Hamori, (2004) , Rabiner (1989).
Decoding:
Similarly, a model estimate that finds the most probable sequence of hidden states given a sequence of observations is the use of the viterbi algorithm. Let

\[ \delta_i(i) = \max P(q_1, q_2, \ldots q_i = s_i, o_1, o_2, \ldots o_i \mid \lambda) \]  

be the maximal probability of state sequences of the length \( t \) that end in state \( i \) and produce the first \( t \) observations for the given model. The variable \( \delta_i(i) \) stores the probability of observing \( o_1, o_2, \ldots o_i \) using the most probable path. The calculation is similar to the forward algorithm, except that the transition probabilities are maximized at each step, instead of summed.

The viterbi algorithm is as follows;

1. Initialization \( \delta_i(i) = \pi_i, b_i(o_1), 1 \leq i \leq N, \phi_i(i) = 0 \)

2. Recursion: \( \delta_i(j) = \max [\delta_{i-1}(i) a_y] b_j(o_t), 2 \leq t \leq T, \leq j \leq N \)

\[ \phi_i(j) = \arg \max [\delta_{i-1}(i) a_y], 2 \leq t \leq T, 1 \leq j \leq N \]

3. Completion: \( q^* = \arg \max [\delta_T(i)] \)

4. Most probable state sequence backtracking: \( q^*_t = \phi_{t+1}(q^*_{t+1}), t = T - 1, T - 2, \ldots, 1 \)

Learning

If we define \( \lambda = (A, B, \pi) \) to denote set of HMM, then the algorithm developed by Baum and Welch for signal processing application (see Rabiner 1989) are applied to estimate the model parameters \( \lambda = (A, B, \pi) \) that best explains the observation. Implementation of the Baum-Welch algorithm works iteratively to improve the likelihood of \( p(O \mid \lambda) \). This iterative process is the training of the model. The Baum-Welch algorithm is calculated as follows;

1. Initialisation:
   Input initial values of \( \lambda \) and calculate \( p(O \mid \lambda) \) using the forward algorithm.

2. Estimate new values of \( \lambda \)
iterate until convergence:

a. calculate \( \gamma_t(i, j) = p(q_t = s_i, q_{t+1} = s_j | O, \lambda) \) for each \( t, i, j \) using the current \( \lambda \)

\[
\gamma_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}
\] (8)

(b) Calculate new \( \lambda \) parameter estimates using \( \gamma_t(i, j) \).

(c) Calculate \( p(O \mid \lambda) \) with new \( \lambda \) values.

3. Goto step 4 if two consecutive calculations of \( p(O \mid \lambda) \) are equal. Else repeat iterations.

4. Output \( \lambda \)

The parameters of the HMMs are estimated by using equation (8). Full details of the Baum-Welch procedure for parameter estimation, as well as the various implementation issues, are described in Rabiner (1989)

2.2 Stochastic volatility model

Stochastic volatility models (see Shephard (1996) for a review) are a variant of the general state space approach presented here. SV model belong to class of Hidden Markov model with non-linear Gaussian state-space model and they take the volatility of the data into account. The SV model due to Taylor (1982) can be expressed as an autoregressive (AR) process:

\[
x_t = \phi x_{t-1} + w_t
\] (9)

\[
r_t = \beta \exp \left( \frac{x_t}{2} \right) v_t
\] (10)

where \( w_t \sim N(0, \tau) \), \( x_0 \sim N(\mu_0, \sigma_0^2) \), \( v_t \sim N(0, 1) \), \( \{r_t\}_{t \geq 0} \) is the log-returns on day \( t \), we call \( \beta \) the constant scaling factor, so that \( \{x_t\}_{t \geq 0} \) represents the log of volatility of the data, \( \log(\sigma_t^2) \) where \( \sigma_t^2 = \text{var}(r_t) \). In order to ensure stationarity of \( r_t \), it is assume that \(|\phi| < 1\). Squaring (10) and taking the logarithm of it results in a linear equation (11),

\[
y_t = \alpha + x_t + z_t
\] (11)
Equations (9) & (11) form the version of the SV model which can be modified in many ways; together they form a linear, non-Gaussian, state-space model for which (11) is the observation equation and (9) is the state equation.

2.2.1 Stochastic Volatility with heavy -tailed distribution

The standard form of the SV model is given in equations (9) & (10). In equation (9) \( v_t \) follows a normal distribution. Various authors have argued that real data may have heavier tails than can be captured by the standard SV model.

An extension of the linearized version of the SV model (see equation (9) and (11), wherein it is assumed that the observational noise process, \( z_t \), is a student-t distribution is considered. The model, first presented in Shumway and Stoffer (2006), retains the state equation for the volatility as:

\[
x_t = \phi x_{t-1} + w_t
\]

but the proposed student-t distribution with degrees of freedom, \( \nu \), for the observation error term, \( z_t \), effects a change in the observation equation:

\[
y_t = \alpha + x_t + z_t \quad z_t \sim t_{\nu}, t = 1, \ldots, n, (12)
\]

For the parameter estimates of the proposed SV model with student-t, the likelihood functions have been maximized by using the Sequential Monte Carlo Expectation Maximization algorithm (Nkemnole et. al., 2011) in the MATLAB optimization routines.

2.3 HMM with stochastic volatility Model

Our model is a blend of the original SV model and HMMs. To start with, we use HMMs to divide the entire time series into regimes with different volatility levels. The return of the time series is assumed to be modelled by a mixture of probability densities and each density function corresponds to a hidden state with its mean and variance. In the HMMs, Viterbi algorithm is employed in finding the state sequence in the time series (Rabiner, 1989). Subsequently we get the subsets of original time series corresponding to different states (volatility levels). Afterwards,
within each regimes, we allow SV model with different parameter sets to model the conditional variance as:

\[ x_t = \phi^i x_{t-1} + w_t, \quad w_t \sim N(0, \tau) \]

\[ y_t = \alpha^i + x_t + z_t, \]

where \( i \) denotes the state of the time series at time \( t \). \( \phi^i, \tau^i, \) and \( \alpha^i \) are the parameter sets of the SV model related to state \( i \).

Thirdly, for the volatility forecast \( \sigma_i^2 \), \( \{x_i\}_{i=0}^{\infty} \) represents the log of volatility of the data, \( \log(\sigma_i^2) \) where \( \sigma_i^2 = \text{var}(r_i) \) of the global model, there is need for us to predict the state \( i \) of time series at time \( t+1 \) (next state). To make the prediction of the next state, we define \( \alpha_i(i) = P(y_1, y_2, \ldots, y_t, q_t = s_i | \lambda) \), we can then estimate the probability of next state in terms of the transition probabilities \( a_{ij} \) as:

\[
P(q_{t+1} = S_j | y_1, y_2, \ldots, y_t, \lambda) = \frac{P(y_1, y_2, \ldots, y_t, q_{t+1} = S_j | \lambda)}{P(y_1, y_2, \ldots, y_t | \lambda)} (13)
\]

\[
= \sum_{j=1}^{N} \alpha_i(i) a_{ij} \sum_{j=1}^{N} \left( \sum_{i=1}^{N} \alpha_i(i) a_{ij} \right) (14)
\]

where \( \alpha_i(i) \) can be estimated from the Baum-Welch algorithm forward-backward algorithm Rabiner, (1989). After the next state \( i \) at time \( t+1 \) has been determined as above, we choose the corresponding SV model with parameter sets \( \phi^i, \tau^i, \) and \( \alpha^i \) to make volatility forecast.

### 3 Volatility Forecast Evaluation and Comparison

#### 3.1 Data and Methodology

Both simulated data sets and real financial data sets were utilized in the volatility forecast experiments. Both the in-sample and the out-of-sample forecasting performances were considered. To start with, we used simulated data set to verify if the proposed model solves the problems of excessive persistence in SV model; we generated more than 11000 observations and
discarded the initial 10000 samples. The diagonal elements \( a_{ii} \) of the transition matrix \( A \) are \( a_{11} = 0.93 \) and \( a_{22} = 0.91 \).

Then we employed the use of real financial data sets (stock return time series) in our experiments to establish the viability of the proposed model. The real financial data sets consist of the daily exchange rate series of the Nigerian Naira, Ghana Cedi, British Pound and Euro, all against the U. S. Dollars (from March 3, 2009 to March 3, 2011) a total of 1911 data set. The total data set is divided into in-sample and out-of-sample data set.

3.1.1 Jarque-Bera Statistics

Jarque-Bera statistics is applied to examine the non-normality of the exchange rate series.

![Figure 1. Naira/dollar exchange rate index summary statistics](image)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Naira/Dollar rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>36.28380</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>64.67169</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.220108</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.489442</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>194.4878</td>
</tr>
<tr>
<td>probability</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

Figure 1 shows a positive skewness, 1.220108, and a high positive kurtosis, 3.489442. With reference to the Jarque-Bera statistics, Naira/dollar exchange rate index is non-normal at the confidence interval of 99%, since probability is 0.000000 which is less than 0.01. Consequently, there is need to convert the Naira/dollar exchange rate index series into the return series.
Figure 2. Cedi/dollar exchange rate index summary statistic

Figure 2 shows a positive skewness, 1.220410 as well as a positive kurtosis, 3.490659. As indicated by Jarque-Bera statistics, the Cedi/dollar exchange rate index is non-normal at the confidence interval of 99%, since probability is 0.0000 which is less than 0.01. So the need also arises to convert the Cedi/dollar exchange rate index series into the return series.

Figure 3. Euro/dollar exchange rate index summary statistics

Figure 3 shows a positive skewness, 1.224487, and a positive kurtosis, 3.506883. As indicated by the Jarque-Bera statistics, Euro/dollar exchange rate index is non-normal at the confidence interval of 99%, since probability is 0.0000 which is less than 0.01; hence the need to convert the Euro/dollar exchange rate index series into the return series.
Table 4. Statistics for Pound/Dollar Exchange Rate Index

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Pound/Dollar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.152918</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.272588</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.220855</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.492430</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>194.6209</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

Figure 4. Pound/dollar exchange rate index summary statistics

Figure 4 shows a positive skewness, 1.220855, and a positive kurtosis, 2.492430. As indicated by the Jarque-Bera statistics, Euro/dollar exchange rate index is non-normal at the confidence interval of 99%, since probability is 0.0000 which is less than 0.01; hence the need to convert the Euro/dollar exchange rate index series into the return series.

3.1.2 Transformation of the exchange rate index series of the Nigerian Naira, Ghana Cedi, British Pound and Euro

On the whole, the movements of the stock indices series are non-stationary, and therefore, not suitable for the study purpose. The stock indices series are transformed into their returns so that we get stationary series. The transformation is:

\[ r_t = 100 \ln \frac{P_t}{P_{t-1}} \]  \hspace{1cm} (15)

where \( r_t \), \( P_t \) is the exchange rate at time index \( t \), \( P_{t-1} \) the exchange rate just prior to the time \( t \).

3.1.3 Augmented Dickey-Fuller (ADF) Test and Phillips-Perron (PP) Test on Naira/Dollar, Cedi/Dollar, Pound/Dollar and Euro/Dollar exchange rates index Returns Series

Both the ADF and PP tests are used to obtain verification regarding whether Naira/Dollar, Cedi/Dollar, Pound/Dollar and Euro/Dollar exchange rates return series is stationary or not.
Table 1. ADF test on Naira/Dollar, Cedi/Dollar, Pound/Dollar and Euro/Dollar exchange rates returns

<table>
<thead>
<tr>
<th>t-Statistic</th>
<th>Naira/Dollar index</th>
<th>Cedi/Dollar index</th>
<th>Pound/Dollar index</th>
<th>Euro/Dollar index</th>
</tr>
</thead>
<tbody>
<tr>
<td>. ADF test statistic</td>
<td>-43.12567</td>
<td>-45.56412</td>
<td>-47.34789</td>
<td>-46.78622</td>
</tr>
<tr>
<td>1% level</td>
<td>-3.331562</td>
<td>-3.33253</td>
<td>-3.331562</td>
<td>-3.33253</td>
</tr>
<tr>
<td>5% level</td>
<td>-2.751341</td>
<td>-2.751341</td>
<td>-2.751341</td>
<td>-2.751341</td>
</tr>
<tr>
<td>Test critical values</td>
<td>-2.456200</td>
<td>-2.456200</td>
<td>-2.456200</td>
<td>-2.456200</td>
</tr>
<tr>
<td>Prob.</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 1 shows that the values of ADF test statistic, -43.12567, is less than its test critical value, -2.751341, at 5%, level of significance which implies that the Naira/Dollar exchange rates return series is stationary. The result of ADF test also demonstrates that the Cedi/Dollar, Pound/Dollar and Euro/Dollar return series are stationary, as the values of ADF test statistic is less than its test critical value.

Table 2. PP test on Naira/Dollar, Cedi/Dollar, Pound/Dollar and Euro/Dollar exchange rates returns

<table>
<thead>
<tr>
<th>t-Statistic</th>
<th>Naira/Dollar index</th>
<th>Cedi/Dollar index</th>
<th>Pound/Dollar index</th>
<th>Euro/Dollar index</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP test statistic</td>
<td>-43.32035</td>
<td>-45.80403</td>
<td>-47.34789</td>
<td>-46.78622</td>
</tr>
<tr>
<td>1% level</td>
<td>-3.331562</td>
<td>-3.33253</td>
<td>-3.331562</td>
<td>-3.33253</td>
</tr>
<tr>
<td>5% level</td>
<td>-2.751341</td>
<td>-2.751341</td>
<td>-2.751341</td>
<td>-2.751341</td>
</tr>
<tr>
<td>Test critical values</td>
<td>-2.456200</td>
<td>-2.456200</td>
<td>-2.456200</td>
<td>-2.456200</td>
</tr>
<tr>
<td>Prob.</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 2 illustrates the results of the PP test and proves that the Naira/Dollar index returns series is stationary, as the values of PP test statistic, -43.32035, is less than its test critical value, -2.751341, at the level of significance of 5%. The outcome of the PP test equally shows that the Cedi/Dollar, Pound/Dollar and Euro/Dollar exchange rates returns series are stationary, since the values of PP test statistic is less than its test critical value.
3.2 Summary Statistics of the Naira/Dollar, Cedi/Dollar, Pound/Dollar and Euro/Dollar exchange rates returns

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Naira/Dollar rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.001385</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.708650</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.074139</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>8.805879</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>735.0376</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

**Figure 5. Naira/dollar exchange rate index returns summary statistics**

Figure 5 reveals a negative skewness, -0.074139, and a positive kurtosis, 8.805879. As indicated by the Jarque-Bera statistics, the Naira/dollar exchange rate index returns series is non-normal at 95% confidence level, since probability is 0.0000 which is less than 0.05.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Cedi/Dollar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.006258</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.536507</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.096923</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>13.11769</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>2312.255</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

**Figure 6. Cedi/dollar exchange rate index returns summary statistics**

Figure 6 also reveals a negative skewness, -0.096923, and a positive kurtosis, 13.11769. Based on the Jarque-Bera statistics, the Cedi/dollar exchange rate index returns series is non-normal at 5% level of significance, because the probability, 0.0000, is less than 0.05.
Figure 7. Euro/dollar exchange rate index returns summary statistics

Figure 7 also reveals a negative skewness, -0.434943, and a positive kurtosis, 7.993814. Based on the Jarque-Bera statistics, the Euro/dollar exchange rate index returns series is non-normal at 5% level of significance, because the probability, 0.0000, is less than 0.05.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Euro/Dollar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.001676</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.488392</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.434943</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.993814</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>559.9343</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

Figure 8. Pound/dollar exchange rate index summary statistics

Figure 8 also reveals a negative skewness, -0.022958, and a positive kurtosis, 4.262290. Based on the Jarque-Bera statistics, the Pound/dollar exchange rate index returns series is non-normal at 5% level of significance, because the probability, 0.0000, is less than 0.05.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Pound/Dollar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.020803</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.506541</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.022958</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.262290</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>34.76827</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000000</td>
</tr>
</tbody>
</table>
4 Empirical Results and Evaluation

As the actual volatility at time \( t \) is not observable, there is need for some measures of volatility to assess the forecasting performance. In this paper we apply the standard approach suggested by Pagan and Schwert, (1990). A proxy for the actual volatility \( \sigma_i^2 \) is given by

\[
\hat{\sigma}_i^2 = (r_i - \bar{r})^2
\]

where \( \bar{r} \) is the mean of the time series over the sample period. The statistical performance measures Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), are applied to select the best performing model both in the in-sample data set and the out-of-sample data set independently in this study:

\[
\text{MSE} = \frac{\sum_{i=1}^{n} (\hat{\sigma}_i^2 - \sigma_i^2)^2}{n}
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{\sigma}_i^2 - \sigma_i^2|
\]

\[
\text{MAPE} = \frac{\sum_{i=1}^{n} |\hat{\sigma}_i^2 - \sigma_i^2|/\sigma_i^2}{n}
\]

where \( \hat{\sigma}_i^2 \) is the forecasted variance and \( \sigma_i^2 \) the actual variance time period \( t \) and \( n \) is the number of forecasts.

4.1 Statistical Performance

The evaluation results are shown in Tables 3 and 4 below. A two-state HMM-SV model was used in our experiments. In both tables, \( t-v \) represents true value, HSV stands for HMM-SV model and SV stands for SV model. \( s_1 \) and \( s_2 \) designate the two states with low and high volatility levels, respectively. \( \text{MSE}_1, \text{MAE}_1 \) and \( \text{MAPE}_1 \) are the in-sample MSE, MAE and MAPE while \( \text{MSE}_2, \text{MAE}_2 \) and \( \text{MAPE}_2 \) are the out-of-sample MSE, MAE and MAPE.

<table>
<thead>
<tr>
<th>Models</th>
<th>( \phi )</th>
<th>( \tau )</th>
<th>( \alpha )</th>
<th>( \text{MSE}_1 )</th>
<th>( \text{MAE}_1 )</th>
<th>( \text{MAPE}_1 )</th>
<th>( \text{MSE}_2 )</th>
<th>( \text{MAE}_2 )</th>
<th>( \text{MAPE}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-v</td>
<td>( s_1 )</td>
<td>0.63</td>
<td>0.8507</td>
<td>2.1662</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( s_2 )</td>
<td>0.70</td>
<td>1.2545</td>
<td>2.0945</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( s_1 )</td>
<td>0.74</td>
<td>0.9408</td>
<td>2.0761</td>
<td>0.2601</td>
<td>0.2102</td>
<td>0.2334</td>
<td>0.1210</td>
<td>0.1716</td>
</tr>
<tr>
<td></td>
<td>( s_2 )</td>
<td>0.70</td>
<td>1.2545</td>
<td>2.0945</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMM-SV</td>
<td>( s_1 )</td>
<td>0.64</td>
<td>1.3654</td>
<td>2.1465</td>
<td>0.1111</td>
<td>0.0753</td>
<td>0.1534</td>
<td>0.0281</td>
<td>0.0484</td>
</tr>
<tr>
<td></td>
<td>( s_2 )</td>
<td>0.78</td>
<td>1.3534</td>
<td>2.1134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Statistical performance results for the stock return data sets and the parameter sets obtained from HMM-SV and SV models

<table>
<thead>
<tr>
<th>Stock Exchange</th>
<th>Models</th>
<th>$\phi$</th>
<th>$\tau$</th>
<th>$\alpha$</th>
<th>MSE$_1$</th>
<th>MAE$_1$</th>
<th>MAPE$_1$</th>
<th>MSE$_2$</th>
<th>MAE$_2$</th>
<th>MAPE$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naira/Dollar</td>
<td>SV</td>
<td>0.8485</td>
<td>4.0273</td>
<td>4.3205</td>
<td>0.3401</td>
<td>0.3211</td>
<td>0.2334</td>
<td>0.2401</td>
<td>0.2211</td>
<td>0.3334</td>
</tr>
<tr>
<td></td>
<td>HMMSV $S_1$</td>
<td>0.7875</td>
<td>3.4771</td>
<td>6.9980</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_2$</td>
<td>0.0685</td>
<td>1.2341</td>
<td>3.8746</td>
<td>0.1021</td>
<td>0.1743</td>
<td>0.3534</td>
<td>0.0021</td>
<td>0.0743</td>
<td>0.3524</td>
</tr>
<tr>
<td>Cedi/Dollar</td>
<td>SV</td>
<td>0.9869</td>
<td>4.1936</td>
<td>5.3824</td>
<td>0.1370</td>
<td>0.1716</td>
<td>0.2265</td>
<td>0.1360</td>
<td>0.1706</td>
<td>0.2255</td>
</tr>
<tr>
<td></td>
<td>HMMSV $S_1$</td>
<td>0.8127</td>
<td>4.2368</td>
<td>4.7144</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_2$</td>
<td>0.0712</td>
<td>3.1134</td>
<td>2.1345</td>
<td>0.0210</td>
<td>0.0595</td>
<td>0.1473</td>
<td>0.0110</td>
<td>0.0495</td>
<td>0.1464</td>
</tr>
<tr>
<td>Pound/Dollar</td>
<td>SV</td>
<td>0.9770</td>
<td>2.1311</td>
<td>0.7654</td>
<td>0.1783</td>
<td>0.0728</td>
<td>0.1922</td>
<td>0.1773</td>
<td>0.0718</td>
<td>0.1812</td>
</tr>
<tr>
<td></td>
<td>HMMSV $S_1$</td>
<td>0.9050</td>
<td>1.3136</td>
<td>0.9883</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_2$</td>
<td>0.0805</td>
<td>1.2136</td>
<td>0.8564</td>
<td>0.0516</td>
<td>0.0983</td>
<td>0.3452</td>
<td>0.0416</td>
<td>0.0783</td>
<td>0.3332</td>
</tr>
<tr>
<td>Euro/Dollar</td>
<td>SV</td>
<td>0.9754</td>
<td>2.3108</td>
<td>0.7627</td>
<td>0.1706</td>
<td>0.2601</td>
<td>0.3988</td>
<td>0.0502</td>
<td>0.1402</td>
<td>0.3678</td>
</tr>
<tr>
<td></td>
<td>HMMSV $S_1$</td>
<td>0.8871</td>
<td>1.4605</td>
<td>1.2590</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_2$</td>
<td>0.0762</td>
<td>1.3605</td>
<td>1.1590</td>
<td>0.0956</td>
<td>0.1943</td>
<td>0.5578</td>
<td>0.0144</td>
<td>0.0943</td>
<td>0.5548</td>
</tr>
</tbody>
</table>

The above results are indicative that that HMM-SV model capture the volatility structure changes processes between two different volatility regimes with different volatility persistence $\phi$. Nonetheless, the SV model cannot capture such volatility structure changes and always show very high volatility persistence. Consequently, HMM-SV model offers better volatility forecasts as the MSE (MAE) of HMM-SV model is considerably smaller than the SV models for most of the time.

5. Conclusion

The volatility persistence of widely-used SV model is usually too high leading to poor volatility forecasts. The root for this excessive persistence seems to be the structure changes (e.g. shift of volatility levels) in the volatility processes, which the SV model cannot capture. As we developed our HMM-SV model to allow for both different volatility states in time series and state specific SV model within each state, the empirical results for both artificial data and real financial data show that the excessive persistence problems disappeared. Accordingly, the results for both in-sample and out-of-sample evaluation forecasting performance confirm that our model outperforms widely-used SV model. Hence, the results suggest that it is promising to deepen the study of volatility persistence, the hidden regime-switching mechanisms inclusive. On long run, this will improve volatility forecasts in future research.
References


