MATIMYÁS MATEMATIKA
ISSN 0115-6926
Journal of the Mathematical Society of the Philippines
Vol. 39 Special Issue (2016) pp. 25-34

A Regime-Switching Lognormal Model of Philippine Stock Exchange Index (PSEi) Returns

Paul Angelo G. Dolores
Institute of Mathematics
University of something
Quezon City, Philippines
pagdolores@gmail.com

Abstract

In this study, we model the returns on the Philippine Stock Exchange Index (PSEi) using a regime-switching lognormal (RSLN) model. Monthly data on the index were obtained from a Technistock [1] terminal, over the period 1987 to 2014. We used maximum likelihood estimation to estimate the parameters of the model. We also presented the results of modeling the returns using a simple lognormal model in comparison with the RSLN model. Finally, we performed Monte Carlo simulations to project expected levels of the PSEi for the immediate future.

Key words: Regime-switching lognormal models, Index returns, Maximum likelihood estimation

1 Introduction

The regime-switching lognormal model switches randomly between a finite number of lognormal processes. In particular, the regime-switching lognormal model with two regimes (denoted: RSLN-2) has the ability to capture stock price movement from a low volatility regime to a high volatility regime and vice versa. This also accounts for extreme movements in stock prices. Hardy [3] introduced the RSLN-2 model for stock returns and used it to model the Toronto Stock Exchange (TSE) 300 and Standard and Poor's (S&P) 500 indices. In this paper, the returns of the Philippine Stock Exchange Index (PSEi) were modeled using the RSLN-2 model. The parameters were estimated using maximum likelihood estimation. Finally, Monte Carlo simulations were performed to obtain short-term projection of PSEi levels.

Monthly data for the Philippine Stock Exchange Index (PSEi) were used to fit the model parameters using maximum likelihood estimation. The data were obtained from a Technistock terminal, over the period January 1987 to December 2014 (equivalent to 336 monthly index prices). Figure 1 shows the historical prices of PSEi.

Figure 2 shows the monthly returns on the PSEi. The mean monthly return was 0.0116, which means that on average, the index level increased by 1.16%. The standard deviation of monthly returns was 0.0881.

[1] Technistock is a financial data provider in the Philippines for brokers, fund managers and other financial companies.

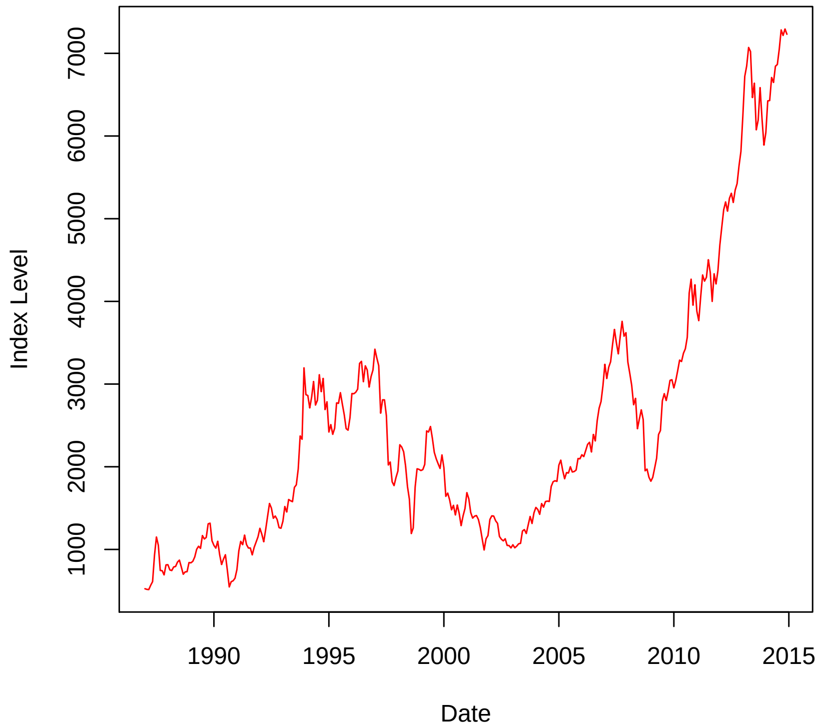


Figure 1: Monthly PSEi Level (January 1987-December 2014)

2 Theoretical Framework

Let S_t be the stock price at any time $t > 0$. The lognormal model assumes that the stock return random variable follows a lognormal distribution with parameters m and v, i.e. for any $k > 0$,

$$\frac{S_{t+k}}{S_t} \sim LN(m = k\mu, v = \sqrt{k}\sigma).$$

This implies that the log-return random variable $Y_t = \log \frac{S_{t+k}}{S_t}$ follows a normal distribution

$$\log \frac{S_{t+k}}{S_t} \sim N(k\mu, v = k\sigma^2),$$

where $k\mu$ and $\sqrt{k}\sigma$ are the mean and standard deviation, respectively, of the normally distributed random variable. Given this equivalence, the log-return random variable Y_t was used for ease of use and familiarity.

Let $\{\rho_t\}_{t\geq 0}$ denote the regime for the interval $[t, t+1)$, with t measured in months. Since this study focuses on a regime switching model with two regimes, we assume that $\rho_t = 1, 2$.

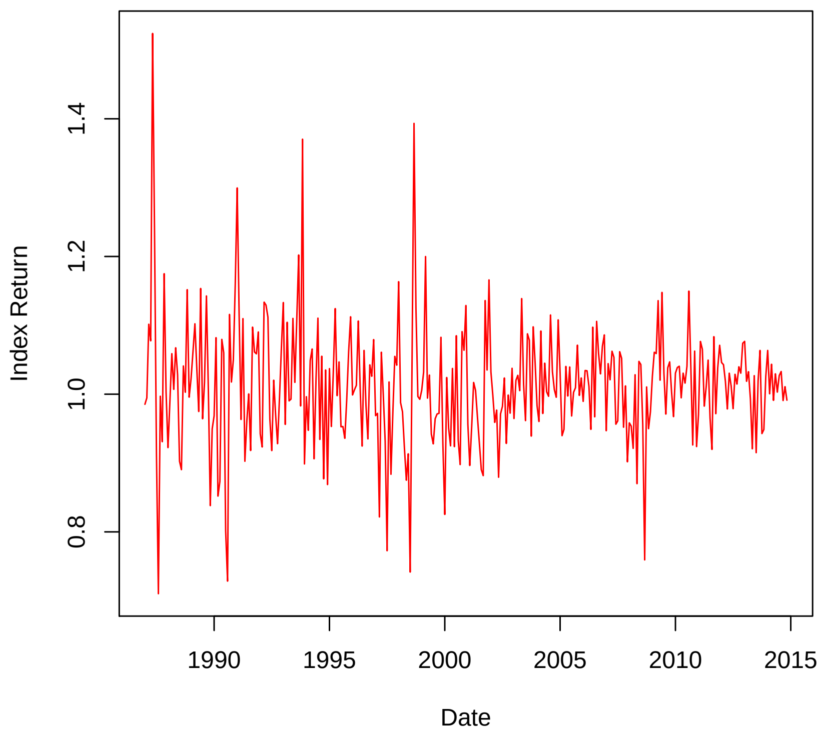


Figure 2: Monthly Total PSEi Returns (January 1987-December 2014)

We define the transition probability $p_{i,j}$ to be the probability that the process transitions to regime j given that the process is in regime i for the previous time period, i.e., $p_{i,j} = \Pr(\rho_{t+1} = j|\rho_t = i)$. Observe that the conditional probability depends only on the regime where the process is currently in and not on the past regimes. Hence the process $\{\rho_t\}_{t\geq 0}$ follows the Markov property and can be called a Markov process. Since we only consider two regimes, we have the transition probabilities $p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}$ and together they form the transition matrix $\mathbf{P}$, written

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}.$$

Under the RSLN-2 model, the stock return process is in regime $\rho_t = 1, 2$ over the interval $[t, t+1)$. Suppose that S_t denotes the index value at time $t > 0$. Then the logarithm of total index return $Y_t = \log S_{t+k}/S_t$ follows a normal distribution with parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$:

$$Y_t = \log \frac{S_{t+1}}{S_t}|\rho_t \sim N(\mu_{\rho_t}, \sigma^2_{\rho_t}),$$

where μ_{ρ_t} and σ_{ρ_t} are the mean and standard deviation, respectively, of the log-return variable in regime $\rho_t = 1, 2$.

We define the regime $\rho_t = 1$ to be the low volatility regime, i.e., the return random variable Y_t is relatively stable over the interval $[t, t+1)$. On the other hand, we define the regime $\rho_t = 2$ to be the high volatility regime, i.e., the return random variable Y_t is relatively unstable over the same interval.

The transition probabilities $p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}$ discussed above take into account the probabilities of changing between the regimes or staying in its current regime.

The six parameters will be estimated, as discussed in the next section, to come up with the desired model: $\mathbf{\Theta} = \{\mu_1, \mu_2, \sigma_1, \sigma_2, p_{1,2}, p_{2,1}\}$.

3 Methods

Monthly data for the Philippine Stock Exchange Index (PSEi) were obtained over the period January 1987 to December 2014 (equivalent to 336 monthly index prices). The maximum likelihood estimation is used to determine the parameter vector $\mathbf{\Theta}$ of the RSLN-2 model.

Let $Y_1, Y_2, \cdots, Y_n$ where $n = 335$, be the log-return random variables with common density function $f(y; \mathbf{\Theta})$. The likelihood function L is given by the joint probability distribution function of the variables, written

$$L(\mathbf{\Theta}) = f(y_1, y_2, \cdots, y_n; \mathbf{\Theta}), \tag{1}$$

which represents the probability of obtaining the sample results $y_1, y_2, \cdots, y_n$ given the value of the parameter vector $\mathbf{\Theta}$ which is initially unknown.

The goal of this method is to obtain an estimate $\widehat{\mathbf{\Theta}}$ of the parameter vector $\mathbf{\Theta}$ that maximizes the value of the likelihood function L. We call $\widehat{\mathbf{\Theta}}$ the maximum likelihood estimator of the parameter vector $\mathbf{\Theta}$. Oftentimes, it is more convenient to find a $\widehat{\mathbf{\Theta}}$ that maximizes the logarithm of the likelihood function, also called the log likelihood function, denoted $l(\mathbf{\Theta}) = \log L(\mathbf{\Theta})$.

As for the RSLN-2 model, the log-return sequence $\{Y_t\}$ exhibits some serial dependence so the conditional distribution function is iteratively used

$$f(y_t | y_{t-1}) = \frac{f(y_{t-1}, y_t)}{f(y_{t-1})}.$$

The modified likelihood function L becomes

$$L(\mathbf{\Theta}) = f(y_1; \mathbf{\Theta}) f(y_2; \mathbf{\Theta} | y_1) f(y_3; \mathbf{\Theta} | y_1, y_2) \cdots, f(y_n; \mathbf{\Theta} | y_1, y_2, \cdots, y_{n-1}) \tag{2}$$

and consequently, the log likelihood function l becomes

$$l(\mathbf{\Theta}) = \log L(\mathbf{\Theta}) = \sum_{i=1}^{n} \log f(y_i; \mathbf{\Theta} | y_1, y_2, \cdots, y_{i-1}). \tag{3}$$

We use the methodology suggested by Hardy [3] in computing for the value of the log likelihood function, which is discussed in detail below. All of the computations illustrated below were done in Microsoft Excel.

The contribution to the log likelihood function of the t-th observation is

$$\log f(y_t; \mathbf{\Theta} | y_1, y_2, \cdots, y_{t-1}).$$

Following the results from Hamilton and Susmel [2], we compute this recursively by calculating for each t

$$f(y_t, \rho_t, \rho_{t-1} | y_{t-1}, y_{t-2}, \cdots, y_1) = f_Y(y_t | \rho_t) \Pr(\rho_t | \rho_{t-1}) \Pr(\rho_{t-1} | y_{t-1}, y_{t-2}, \cdots, y_1) \tag{4}$$

where:

- $f_Y(y_t|\rho_t)$ is the density function of the normal random variable Y_t, given by

$$f_Y(y_t|\rho_t) = \frac{1}{\sigma_{\rho_t}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_t-\mu_{\rho_t}}{\sigma_{\rho_t}}\right)^2} \tag{5}$$

- $\Pr(\rho_t|\rho_{t-1})$ is the transition probability between regimes from time $t-1$ to time t and is equal to $p_{i,j}$ for $i, j = 1, 2$.

- $\Pr(\rho_{t-1}|y_{t-1}, y_{t-2}, \cdots, y_1)$ is a probability function obtained through recursion. The starting values for this recursion are

$$\Pr(\rho_1 = 1|y_1) = \frac{f(\rho_1 = 1, y_1)}{f(y_1)} \text{ and } \Pr(\rho_1 = 2|y_1) = \frac{f(\rho_1 = 2, y_1)}{f(y_1)}$$

where

$$\begin{aligned} f(\rho_1 = 1, y_1) &= \pi_1 f_Y(y_1|\rho_1 = 1) \\ f(\rho_1 = 2, y_1) &= \pi_2 f_Y(y_1|\rho_1 = 2) \\ f(y_1) &= f(\rho_1 = 1, y_1) + f(\rho_1 = 2, y_1) \end{aligned}$$

The constants π_1 and π_2 were obtained from the invariant distribution $\boldsymbol{\pi}$ for the process, where $\boldsymbol{\pi} = (\pi_1\ \pi_2)$. Given no information about the historical movements of the process, the probability that the process is in regime 1 is equal to π_1 and similarly, the probability that the process is in regime 2 is equal to π_2. Hence the transition probabilities, as defined by the transition matrix $\mathbf{P}$, return the same distribution under $\boldsymbol{\pi}$, i.e.,

$$\begin{aligned} \boldsymbol{\pi}\mathbf{P} &= \boldsymbol{\pi} \\ \Leftrightarrow (\pi_1\ \pi_2)\begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix} &= (\pi_1\ \pi_2) \end{aligned} \tag{6}$$

Then using (6) and the fact that $\pi_1 + \pi_2 = 1$, we obtain the values for π_1 and π_2:

$$\begin{aligned} \pi_1 &= \frac{p_{2,1}}{p_{1,2} + p_{2,1}} \\ \pi_2 &= 1 - \pi_1 \end{aligned}$$

For the succeeding values of the recursion,

$$\Pr(\rho_{t-1}|y_{t-1}, y_{t-2}, \cdots, y_1) = \sum_{\rho_{t-2}=1}^{2} \frac{f(y_{t-1}, \rho_{t-1}, \rho_{t-2}|y_{t-2}, y_{t-3}, \cdots, y_1)}{f(y_{t-1}|y_{t-2}, y_{t-3}, \cdots, y_1)} \tag{7}$$

Now, summing the values of (4) for $\rho_t, \rho_{t-1} = 1, 2$ yields $f(y_t|y_{t-1}, y_{t-2}, \cdots, y_1)$. Then, the contribution to the log likelihood of the t-th observation is then obtained by taking the logarithm of $f(y_t|y_{t-1}, y_{t-2}, \cdots, y_1)$. Finally, we obtain the log likelihood function by calculating the sum

$$\sum_{t=1}^{n} \log f(y_t|y_{t-1}, y_{t-2}, \cdots, y_1).$$

After obtaining the value of the log likelihood function l, its value is maximized to determine the maximum likelihood estimate $\widehat{\boldsymbol{\Theta}}$ of the parameter vector $\boldsymbol{\Theta}$. This optimization problem is specified as

$$\text{Maximize } \sum_{t=1}^{n} \log f(y_t|y_{t-1}, y_{t-2}, \cdots, y_1)$$

subject to

$$\sigma_1, \sigma_2 \geq 0$$

$$0 \leq p_{i,j} \leq 1, \text{ for } i, j = 1, 2$$

$$p_{1,1} + p_{1,2} = 1$$

$$p_{2,1} + p_{2,2} = 1$$

The results are discussed in the next section.

4 Results and Discussion

The optimization problem was solved using Microsoft Excel's Solver[2] tool and the results are shown in Table 1.

Table 1: Maximum Likelihood Estimation Results for RSLN-2 Model

$\hat{\mu}_1 = 0.01089$	$\hat{\sigma}_1 = 0.06199$	$\hat{p}_{1,2} = 0.03847$
$\hat{\mu}_2 = -0.00725$	$\hat{\sigma}_2 = 0.15801$	$\hat{p}_{2,1} = 0.19231$
loglikelihood=378.2216		

Under the low volatility regime, or regime 1, the log return process had a monthly mean of 0.01089 and monthly standard deviation of 0.06199. On the other hand, in the high volatility regime, or regime 2, the process had a monthly mean of -0.00725 and indeed had higher standard deviation than that of regime 1 at 0.15801. Regime 2 can be associated with periods of falling index levels, as described by the negative mean. The estimated probabilities indicated that the stock return process was more likely to move from regime 2 to regime 1 than the other way around.

The maximum likelihood estimates for the lognormal model were given by $\widehat{\mathbf{\Theta}} = (\hat{\mu}, \hat{\sigma})$. These were obtained by maximizing the log likelihood function for the lognormal model, defined as

$$l\left(\widehat{\mathbf{\Theta}}\right) = \sum_{t=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left\{-\frac{1}{2}\left(\frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}\right)^2\right\}\right). \tag{8}$$

As with the RSLN-2 model, equation (8) was maximized using Microsoft Excel's Solver tool. Table 2 shows the values obtained for the maximum likelihood estimates and the log likelihood.

Table 2: Maximum Likelihood Estimation Results for Lognormal Model

$\hat{\mu} = 0.00783$
$\hat{\sigma} = 0.08632$
loglikelihood=345.3018

We used two model selection criteria to compare the lognormal and RSLN-2 models, namely: Schwartz-Bayes criterion and Akaike information criterion. These aimed to determine the effect of increasing the number of model parameters on the fit of the model.

[2]Solver is an add-in for Microsoft Excel that can be used for numerically solving optimization problems.

The Schwartz-Bayes criterion (SBC) and Akaike information criterion (AIC) are measures of the quality of fit of a model (log likelihood value) to the given data set, with consideration to the number of model parameters. The values of SBC and AIC are given by

$$\text{AIC} = l_j - k_j \tag{9}$$

$$\text{SBC} = l_j - \frac{1}{2}k_j \ln n \tag{10}$$

where

l_j : maximized value of the log likelihood function of model j
k_j : number of parameters of model j
n : sample size, equal to 335 historical log-returns

The model with higher AIC and SBC values is preferred. Table 3 summarizes the results for these two criteria.

Table 3: Comparison of Lognormal and RSLN-2 Models

Model	No. of Parameters	Loglikelihood	SBC	AIC
RSLN-2	6	378.22	360.78	372.22
Lognormal	2	345.30	339.49	343.30

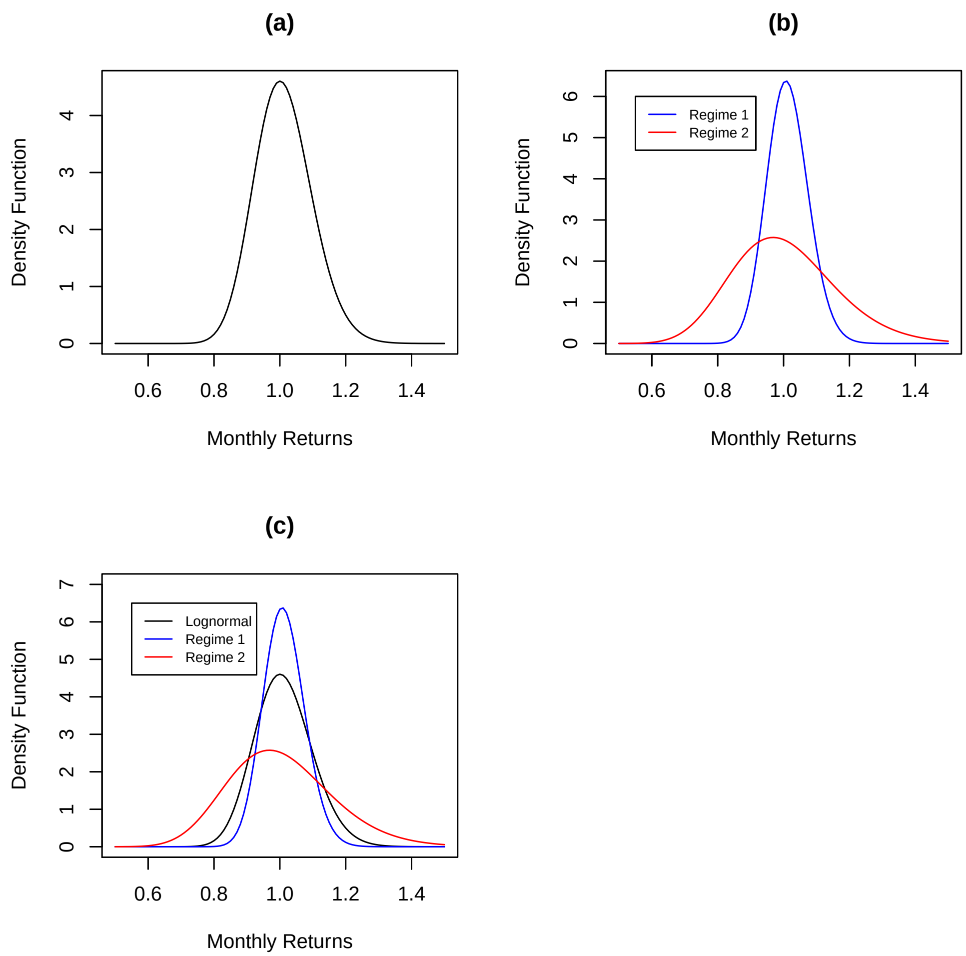
This comparison was done to determine if using a model with more parameters provided a significant improvement in fit compared to the simpler model. The argument against using more parameters is the principle of parsimony. It may appear that increasing the number of parameters would increase the goodness of fit but by doing so, we might over-fit the model to the data. The log likelihood values showed that RSLN-2 provided a better fit than the lognormal model. As discussed above, the two criteria penalized the log likelihood by subtracting a quantity proportional to the number of parameters. Despite the penalty, RSLN-2 still emerged on top over the lognormal model. Figure 3 compares the density functions of the two models graphically.

After estimating the model parameters, we performed Monte Carlo simulations with 10,000 runs to project the level of PSEi over the next two years. The simulation process described below was implemented in R[3].

1. Generate a uniformly distributed random number u on the interval $[0, 1]$.
2. Suppose that $\Pr(\rho_0 = 1) = \pi_1$. If $u < \pi_1$, assume $\rho_0 = 1$; otherwise, assume $\rho_0 = 2$.
3. Generate $z \sim N(0, 1)$ using u via the Inverse Transform method.
4. Simulate a log-return result using the equation $y_t = \mu_{\rho_{t-1}} + \sigma_{\rho_{t-1}} z$.
5. The return at any time t is given by e^{y_t} and consequently, the index level at time t is $S_t = S_{t-1}e^{y_t}$.
6. Generate a new uniformly distributed random number u on the interval $[0, 1]$.
7. If $u < p_{\rho_0,1}$, then assume $\rho_1 = 1$; otherwise, assume $\rho_1 = 2$.

[3] R is a statistical computing system with robust programming language and capacity for data manipulation and graphical display.

Figure 3: (a) Density function for the simple lognormal model; (b) Density functions for the two regimes of the RSLN-2 model; and (c) Overlay of the density functions



8. Repeat steps (3) to (7) for the required number of months in the projection period.

9. Repeat steps (1) to (8) for the required number of runs/simulations.

The expected value of the 10,000 simulations was computed, which we used as the model's projections. Over the short-term, we measured the model's performance relative to the actual index data. We show the projected index levels for the entire year 2015 and compare it with actual available monthly data until April 2015. Figure 4 and Table 4 show the comparisons.

We also plotted upper and lower bounds corresponding to 95% confidence level to illustrate a degree of uncertainty in the projections. It can be seen that from December 2014's closing price of 7,230.57, the index rose to 7,689.91 in January 2015. On the other hand, the model projected that the index would only grow to 7,304.79. The index continued its rise until March 2015 when it closed at 7,940.49, way above the model's projected value of

Table 4: Projected Monthly PSEi Levels for 2015

Date	Actual PSEi Level	Projected PSEi Level	% Actual / Projected
12/29/2014	7,230.57		
01/31/2015	7,689.91	7,304.79	5.05%
02/27/2015	7,730.57	7,376.74	5.27%
03/31/2015	7,940.49	7,444.74	6.66%
04/30/2015	7,714.82	7,518.95	2.61%
05/29/2015		7,600.03	
06/30/2015		7,675.55	
07/31/2015		7,762.48	
08/28/2015		7,836.37	
09/30/2015		7,913.33	
10/30/2015		7,990.96	
11/27/2015		8,078.20	
12/29/2015		8,164.93	

7,444.74. At the rate the stock index was rising, it seemed that the model underestimated the future growth but we should take note that stock prices could go down anytime. On April 2015, the index dipped to 7,714.82. It was the closest the actual index data had been to the model projection at 7,518.95.

5 Conclusion

The regime-switching lognormal model is able to capture extreme price movements. The low and high volatility regimes adequately describe how the stock market behaves. It provides a better fit to the Philippine Stock Exchange Index data than the simple lognormal model, as evidenced by the log likelihood, Schwartz-Bayes criterion (SBC) and Akaike information criterion (AIC) values.

An improvement to this study is the use of additional goodness-of-fit tests. In [3], the RSLN-2 model was used in option pricing and in evaluating risk measures for equity-linked contracts. Another thing that could be done is to increase the number of simulations, perhaps to hundreds of thousands or even millions. The most important thing to do with these projections is to use them with caution. Ultimately, what matters is the decision we make after careful examination of the model and the projection results.

References

[1] Hamilton, J.D. “A new approach to the economic analysis of non-stationary time series.” Econometrica, 57 (1989): 357–384.

[2] Hamilton, J.D. and Susmel, R. “Autoregressive conditional heteroskedasticity and changes in regime.” Journal of Econometrics, 64 (1994): 307–333.

[3] Hardy, M.R. “A regime-switching model of long-term stock returns.” North American Actuarial Journal, 5(2) (2001): 41–53.

[4] Hardy, M.R. Investment Guarantees. Hoboken, New Jersey: John Wiley & Sons, Inc., 2003.

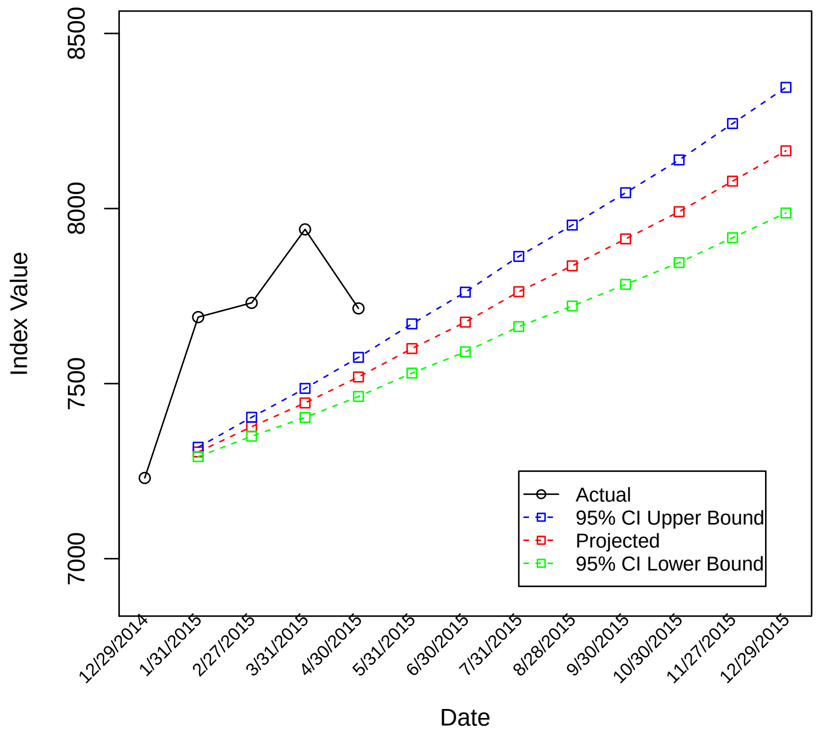


Figure 4: Monthly Actual and Projected PSEi Level (December 2014-December 2015)

[5] International Actuarial Association. Stochastic Modeling. Ontario, Canada, 2010.

[6] Mizrach, B. and Watkins, J. "A Markov switching cookbook." Working Papers, Department of Economics, Rutgers University, (1998).