Developed A Hybrid Sliding Window and GARCH Model for Forecasting of Crude Palm Oil Prices in Malaysia

Amiratul Liyana Mohamad Hanapi, Mahmod Othman, Rajalingam Sokkalingam, Hamzah Sakidin

Department of Fundamental and Applied Science, Universiti Teknologi Petronas

Email: amiratulliyana13@gmail.com, mahmod.othman@utp.edu.my, raja.sokkalingam@utp.edu.my, hamzah.sakidin@utp.edu.my

Abstract. The increase of crude palm oil prices can significantly affect the worldwide economic activities. Therefore, an accurate model to forecast the crude palm oil prices is crucial so that necessary precautionary steps can be taken. In this study, a hybrid between Sliding Window and GARCH model was proposed to improve the forecasting accuracy of crude palm oil prices series. In this model, sliding window partitions is used to aggregate / cluster the original series into several number of constitutive series while GARCH model is utilized to forecast prices based on the selected window to complete variance calculation. A dataset of crude palm oil prices from Malaysian Palm Oil Board was used to test the performance of the proposed model. Direct application of GARCH model was used as a benchmark for effectiveness measurement with the proposed model by comparing mean percentage of absolute error and mean square error. The result has shown that the proposed hybrid sliding window and GARCH model demonstrates better forecasting performance than single GARCH model.

1. Introduction

Malaysia is one of the biggest producers and exporters of crude palm oil (CPO) and palm oil products. The palm oil industry contributes significantly to Malaysia’s export revenue. Therefore, accurate forecasting model to forecast CPO prices is important for a stable and rapid economic development. Various forecasting models are available to forecast the CPO prices. The most common time series forecasting model is ARIMA. Muda et al. [1] who used ARIMA model in their study found that the model produces likely forecasts, which are more accurate and the predicted CPO productions are increasing for the next three months. However, ARIMA model does not reflect recent changes nor incorporate new information [2]. He suggests using ARCH or GARCH model for forecasting nonlinearity.

A nonlinear time series data was analyzed using GARCH model that was introduced by Bollerslev [3]. The GARCH model differentiates several assumptions on volatility effects by treating the weights of each parameter’s estimated value. The model gives new information of weight to long-run average variance from historical data thus allows for both longer memory and more flexible lag structure. The GARCH model is presented as in equation below,

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1} + \beta \sigma_{n-1}^2.$$
where $\gamma$, $\alpha$, and $\beta$ are the weights assigned to $V_n, u_{n-1}$, and $\sigma^2_{n-1}$, respectively. However, the calculated long run average variance does not reflect the influence of each variance as the calculation of long run average variance is derived from the whole series [4].

Many researchers are combining mathematical and statistical models method into GARCH model to seek improvement on the latter. For example, some hybrid models of Artificial Neural Network and GARCH model are better and more efficient than the standalone GARCH model as suggested in a few studies [5][6]. The hybrid models, although more efficient they are, however, complicated in nature as ANN were not easily understood. Another hybrid model named GJR-GARCH model also produces accurate results than the standalone GARCH model. The GJR-GARCH model shows limitation when the model does not simulate stock fluctuations with volatility clustering [7].

Besides that, Kapoor and Bedi [7] improved the time series forecasting model using sliding window algorithm. They considered a span of two adjacent week of previous year’s variation to determine the dependency of current year’s variation. The result exhibited high accuracy in the weather prediction, excluding the highly unpredictable conditions in certain seasonal months. The main task performed by sliding window is to identify a set of cut off points to partition the range into smaller number of intervals [8]. Area partitioning also worked well with the sets of sliding windows in a study of error detection method for ocean depth series data [9].

Sliding window is a common technique used in application of series data and forecasting. This study is an attempt to propose a hybrid sliding window and GARCH model by improving the GARCH model for forecasting crude palm oil prices in Malaysia. Monthly data set of Malaysian crude palm oil prices for the period of January 2009 to December 2017 obtained from the website http://bepi.mpob.gov.my/index.php/en/statistics/price/monthly.html was retrieved for data analysis.

2. Methodology
This study attempts to propose a hybrid model combining sliding window and GARCH model to overcome limitations of GARCH model. The first subsection explains the steps of hybrid sliding window and GARCH model while second subsection explains the evaluation of the proposed model.

2.1 A Hybrid Sliding Window and GARCH Model
The proposed hybrid model composes of four steps as shown in figure 1. Sliding window technique is applied in the step 3, during which the long run variance is replaced by window variance in order to obtain the weights for forecasting. This incorporates more recent returns that provide greater weight [4].
In the first step of the proposed model, the parameters $\gamma$, $\alpha$ and $\beta$ are estimated from the sample of historical data. The approach used is the Maximize Likelihood Method (MLM). It involves choosing values for the parameters that have maximum chance or likelihood of the data [4]. Maximize likelihood is an iterative searching tool to find the parameters in the model that maximize the expression. The sum of $\gamma$, $\alpha$ and $\beta$ is equal to 1.

After we found the optimal parameters, returns of the observations are calculated by squaring the change in rate $S_i$ at the end of month $t$ as shown in equation (1).

$$\delta_i = (S_i - S_{i-1})/S_{i-1},$$  \hspace{1cm} (1)

The third step of the proposed model starts with selection of sliding window for the weight. The selection of sliding window to be the weight of model is determined by Euclidean distance. A matrix size of 12x1 for current year, CY and a matrix size of 24x1 for previous observations, PY that consists of 24 months from previous two years are constructed. The matrix $PY$ is then used to construct 13 slices of sliding windows. The sliding windows are labeled as $SW_i$ with matrix size of 12x1 that are the same size as the matrix $CY$. The partitions of the windows are given as equation (2).

$$SW_i = \{X_{n-i}, X_{n-(i-1)}, X_{n-(i-2)}, X_{n-(i-3)}, \ldots\},$$  \hspace{1cm} (2)

Where $X_n$ is the first observation of current year and $i = 1, 2, \ldots, 13$. Euclidean distance, $Edi_i$, is derived from equation (3).

$$Edi_i = \sqrt{(CY_i - SW_{i,j})^2},$$  \hspace{1cm} (3)

where $j = 1, 2, \ldots, 12$ are the elements in the matrices of $CY$ and $SW$. Then the mean of Euclidean distance of each $Edi$ is calculated. The window that produces the smallest mean of Euclidean distance is selected as the size of forecasting weight. Thus, the total of window size weights is calculated using equation (4) and then the weights are normalized using equation (5).

$$TL = w_n + w_{n-1} + w_{n-2} + \cdots + w_1,$$  \hspace{1cm} (4)

$$W_i = \frac{w_i}{TL},$$  \hspace{1cm} (5)
where \( w_n \) is weight of day \( n \) and \( w_i \) is the weight for each observation. Then, each weight is multiplied with the return of the observation. The multiplication of the weights and returns is summed up to obtain the window variance as shown in equation (6).

\[
V_w = (u_1^2 \times W_1) + (u_{t-1}^2 \times W_2) + (u_{t-2}^2 \times W_3) + \cdots + (u_{t-(n-1)}^2 \times W_n),
\]

where \( V_w \) is the result of window variance, \( u_i^2 \) is return and \( W_i \) is window weight subject to \( \sum_{i=1}^{n} W_i = 1 \). Next, the window variance is used in computation of variance of the proposed model, \( \sigma^2 \), with recursive procedure approach as shown in equation (7). The variance \( \sigma^2 \) is based on the most recent observation of return.

\[
\sigma_n^2 = (V_w \times \gamma) + (u_{n-1}^2 \times \alpha) + (\sigma_{n-1}^2 \times \beta),
\]

where \( u_{n-1}^2 \) is recent return and \( \sigma_{n-1}^2 \) is recent variance. The return \( u_i^2 \) is defined as a continuously compounded return during time \( i \), or can be understood as the return of gain or loss in a particular period. As we can see in the equation (7) which is an enhancement from GARCH model, the long run average variance in GARCH model is replaced with window variance.

In the final step of the proposed model, the future values are obtained using expected forecast variance. The expected forecast variance is computed as in equation (8) by employing window variance as the weight of forecasting.

\[
E[\sigma_{n+t}^2] = V_w + (\alpha + \beta)^t + (\sigma_n^2 - V_w),
\]

where \( n \) represents the month of the calculated variance and \( t \) represents the additional time that will reflect the forecasted month. Finally, the forecasted value is computed as in equation (9).

\[
X_{n+1} = X_n + (X_n \times E[\sigma_n^2]),
\]

where \( X_n \) is the current observation.

2.2 Evaluation of the Proposed Hybrid Model
The mean squared error (MSE) and mean absolute percentage error (MAPE) are applied for comparing and assessing accuracy performance of the model. The MSE and MAPE are shown in equation (10) and equation (11) respectively.

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (\tilde{Y}_i - Y_i)^2,
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\tilde{Y}_i - Y_i|}{Y_i} \times 100,
\]

If the MAPE calculated value is less than 10%, it is interpreted as highly accurate forecasting, between 10–20% is good forecasting, between 20–50% is reasonable forecasting, and over 50% is inaccurate forecasting [10]. Benchmark model for performance comparison is GARCH.

3. Result and Discussion
This section consists of the forecasting of crude palm oil prices by the proposed hybrid model and the GARCH model as well as the comparison of errors between the two models.
3.1 Forecasting Crude Palm Oil Prices by the Proposed Hybrid Model
A monthly data on crude palm oil (CPO) prices (RM) covering January 2009 to December 2017 is obtained from the official website of the Malaysian Palm Oil Board (MPOB). Our strategy for modelling the conditional variance is firstly to estimate the parameters $\gamma$, $\alpha$ and $\beta$ that are respectively assigned to window variance, return, and recent variance as in equation (5). We are interested in choosing the parameters to maximize the sum of the likelihood. This involves an iterative search procedure. The likelihood values were determined in the final iteration of search for the optimal $\gamma$, $\alpha$ and $\beta$. In this dataset, the optimal values of the parameters are $\gamma = 0.68390$, $\alpha = 0.01588$, $\beta = 0.30022$. Subsequently, the returns of observations are computed.

The third step of the model continued with the proposed method. The size of the forecasting was determined from the selected window that produced minimum mean of Euclidean distance. Before the determination of size, the measurements of $CY$ and $PY$ were constructed with the 13 partitioned windows. Some of these measurements were tabulated as in table 1.

Table 1. Some of the observations of current year and previous years with 13 windows.

| Time  | $CY$  | $PY$-W1 | $PY$-W2 | $PY$-W3 | $PY$-W4 | $PY$-W5 | $PY$-W6 | $PY$-W7 |
|-------|-------|---------|---------|---------|---------|---------|---------|---------|
| Jan-17| 3255.00 | 3194.50 | 2919.50 | 2742.50 | 2873.50 | 2628.50 | 2298.00 | 2533.00 |
| Feb-17| 3262.00 | 3255.00 | 3194.50 | 2919.50 | 2742.50 | 2873.50 | 2628.50 | 2298.00 |
| Mar-17| 2965.50 | 3262.00 | 3255.00 | 3194.50 | 2919.50 | 2742.50 | 2873.50 | 2628.50 |
| Apr-17| 2667.50 | 2965.50 | 3262.00 | 3255.00 | 3194.50 | 2919.50 | 2742.50 | 2873.50 |
| May-17| 2784.00 | 2667.50 | 2965.50 | 3262.00 | 3255.00 | 3194.50 | 2919.50 | 2742.50 |
| Jun-17| 2721.00 | 2784.00 | 2667.50 | 2965.50 | 3262.00 | 3255.00 | 3194.50 | 2919.50 |
| Jul-17| 2607.50 | 2721.00 | 2784.00 | 2667.50 | 2965.50 | 3262.00 | 3255.00 | 3194.50 |
| Aug-17| 2610.50 | 2607.50 | 2721.00 | 2784.00 | 2667.50 | 2965.50 | 3262.00 | 3255.00 |
| Sep-17| 2766.50 | 2610.50 | 2607.50 | 2721.00 | 2784.00 | 2667.50 | 2965.50 | 3262.00 |
| Oct-17| 2712.50 | 2766.50 | 2610.50 | 2607.50 | 2721.00 | 2784.00 | 2667.50 | 2965.50 |
| Nov-17| 2652.00 | 2712.50 | 2766.50 | 2610.50 | 2607.50 | 2721.00 | 2784.00 | 2667.50 |
| Dec-17| 2394.50 | 2652.00 | 2712.50 | 2766.50 | 2610.50 | 2607.50 | 2721.00 | 2784.00 |
For example, the calculation of Euclidean distance of Window 1 is shown below.

\[ Edi_1 = \frac{\sum_{i=1}^{95} \sqrt{(2394.50 - 2652.00)^2 + (2652.00 - 2712.50)^2 + (2712.50 - 2766.50)^2 + (2766.50 - 2610.50)^2 + \cdots}}{95} \]

\[ Edi_1 = 126.37. \]

The Euclidean distance for each window is also shown in Table 2. From the table, we can see that Window 1 produced the minimum Euclidean distance. Therefore, the total of window size weights (TL) is 1 and \( W_i = w_i \).

### Table 2. Euclidean distances of 13 windows.

| Window | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Edi    | 126.37 | 197.10 | 248.98 | 297.14 | 338.31 | 367.33 | 394.13 | 412.57 | 432.28 | 442.58 | 446.80 | 465.17 | 489.57 |

The window variance \( V_w \) and variance of the observation \( \sigma^2 \) were then computed using equation (4) and equation (5), respectively. The variance was used in the calculation of expected variance for future measurements. The forecasting was done and table 3 shows some of the data with forecasted values along with their respective errors.

### Table 3. Result of the proposed model.

| Time   | CPO Price | Window Variance | Variance of Proposed Model | Forecast Variance | Forecast Value | Squared Error | | Percentage Error |
|--------|-----------|-----------------|----------------------------|-------------------|---------------|---------------|----------------|------------------|
| Jan-17 | 3255.00   | 0.0004          | 0.0026                     | 0.0011            | 3282.4069     | 751.1366      | 0.8420          |
| Feb-17 | 3262.00   | 0.0000          | 0.0008                     | 0.0003            | 3265.5018     | 12.2623       | 0.1073          |
| Mar-17 | 2965.50   | 0.0083          | 0.0059                     | 0.0075            | 2966.2550     | 0.5701        | 0.0255          |
| Apr-17 | 2667.50   | 0.0101          | 0.0088                     | 0.0097            | 2687.5380     | 401.5225      | 0.7512          |
| May-17 | 2784.00   | 0.0019          | 0.0041                     | 0.0026            | 2810.9753     | 727.6664      | 0.9689          |
| Jun-17 | 2721.00   | 0.0005          | 0.0016                     | 0.0009            | 2728.0831     | 5.0330        | 0.0255          |
| Jul-17 | 2607.00   | 0.0017          | 0.0017                     | 0.0017            | 2609.7434     | 5.0330        | 0.0255          |
| Aug-17 | 2610.50   | 0.0000          | 0.0005                     | 0.0002            | 2614.9948     | 20.2033       | 0.1722          |
| Sep-17 | 2766.50   | 0.0036          | 0.0026                     | 0.0033            | 2766.9692     | 0.2202        | 0.0170          |
| Oct-17 | 2712.50   | 0.0004          | 0.0011                     | 0.0006            | 2721.3561     | 78.4309       | 0.3265          |
| Nov-17 | 2652.00   | 0.0005          | 0.0007                     | 0.0006            | 2653.6120     | 2.5985        | 0.0608          |
| Dec-17 | 2394.50   | 0.0094          | 0.0067                     | 0.0086            | 2395.8264     | 1.7593        | 0.0554          |

The optimum values of \( \gamma = 0.68390, \alpha = 0.01588, \beta = 0.30022 \) calculated by likelihood method used to compute the variance of the proposed model (Table 3) based on equation (7) showed that the absolute percentage error is in between 0% to 1%. The result supports that the estimated parameters are optimum enough and the proposed model produces almost accurate forecast values as the actual CPO prices.
3.2 Forecasting Crude Palm Oil Prices by the GARCH Model

In order to compare the accuracy performance of the proposed hybrid model, the CPO prices data were also analyzed using GARCH model. Table 4 shows the result of GARCH model where the optimal parameters are $\gamma = 0.1464$, $\alpha = 0.0047$, $\beta = 0.8489$ and the long run variance is $\nu_L = 0.0768$.

Table 4. Result of the GARCH model.

| Time   | CPO Price | Variance of GARCH Model | Forecasted variance | Forecasted value | Squared Error | Percentage Error |
|--------|-----------|-------------------------|---------------------|------------------|---------------|-----------------|
| Jan-17 | 3255.00   | 0.0188                  | 0.0273              | 3332.8433        | 6059.5752     | 2.3915          |
| Feb-17 | 3262.00   | 0.0116                  | 0.0212              | 3351.1596        | 7949.4364     | 2.7333          |
| Mar-17 | 2965.50   | 0.0113                  | 0.0209              | 3028.3092        | 3944.9968     | 2.1180          |
| Apr-17 | 2667.50   | 0.0183                  | 0.0269              | 2723.2354        | 3106.4394     | 2.0894          |
| May-17 | 2784.00   | 0.0199                  | 0.0282              | 2858.8229        | 5598.4648     | 2.6876          |
| Jun-17 | 2721.00   | 0.0130                  | 0.0223              | 2797.8269        | 5902.3755     | 2.8235          |
| Jul-17 | 2607.50   | 0.0117                  | 0.0213              | 2665.6640        | 3383.0546     | 2.2306          |
| Aug-17 | 2610.50   | 0.0128                  | 0.0222              | 2666.0183        | 3082.2803     | 2.1267          |
| Sep-17 | 2766.50   | 0.0113                  | 0.0209              | 2827.7837        | 3755.6894     | 2.2152          |
| Oct-17 | 2712.50   | 0.0143                  | 0.0235              | 2769.1817        | 3212.8146     | 2.0896          |
| Nov-17 | 2652.00   | 0.0116                  | 0.0212              | 2714.2614        | 3876.4774     | 2.3477          |
| Dec-17 | 2394.50   | 0.0117                  | 0.0213              | 2445.2104        | 2571.5453     | 2.1178          |

The percentage errors produced by GARCH model were also very low and most of them are within 1% and 4%. The GARCH model produced almost accurate forecast values as the actual CPO prices. From the results, we can say that the parameters of the GARCH model were also optimum enough.

3.3 Comparison of the Proposed Hybrid Model with GARCH Model

From section 3.1 and section 3.2, the proposed model produced 0% to 1% percentage errors while the GARCH model produced 1% to 4% percentage errors. Thus, we can prove that the proposed model performed better than the GARCH model in forecasting CPO prices. In other perspective, the means of the square error and percentage absolute errors are summarized in Table 5. From table 5, the MSE and MAPE of the proposed model are lower than the MSE and MAPE of the GARCH model. This result proved that the proposed model is able to produce better accuracy performance compared to GARCH model.

Table 5. Comparison result of the CPO prices data.

| Model | Error | MSE    | MAPE   |
|-------|-------|--------|--------|
| Proposed | 884.5260 | 0.5369     |
| GARCH   | 4744.7160 | 2.4718     |

4. Conclusion and Future Work

Based on the result and discussion, the proposed hybrid sliding window and GARCH model produced a highly accurate forecasting CPO prices thus giving an advantage in obtaining valuable information pertaining to forecast CPO prices which helps for a stable and rapid economy. Through the error analysis, the research found that the percentage error between actual and forecast value of the proposed model was near 0% to 1% with optimal parameters $\gamma = 0.68390$, $\alpha = 0.01588$, and $\beta = 0.30022$, while the percentage error between actual and forecast value of the GARCH model was near 1% to 4% with its optimal parameters. Both models produced high accuracy results but the proposed model outperformed the GARCH model by producing lower percentage error. Furthermore, the MSE and MAPE of the proposed
model were 884.5260 and 0.5369%, respectively, were significantly lower than the MSE and MAPE of the GARCH model. Therefore, the results proved that the proposed model is able to produce better accuracy performance compared to GARCH model.

For future work, modification on the estimation of parameters can be researched because the starting value of parameter estimation using likelihood method is very sensitive and affects the optimum value of the estimation. In addition, optimization is also key to give the best performance of a complex and complicated application.

Acknowledgement
My appreciation goes to Universiti Teknologi PETRONAS for providing financial support under Grant STIRF entitled “Forecasting the Price of Crude Oil Based on Fuzzy Time Series Model with Combination of Sliding Window Method (Cost Centre: 0153AA-F33)

References
[1] N. Muda, R. Sulong and Z. Mohamed, "Malaysia Crude Oil Production Estimation: an Application of ARIMA Model," in International Conference on Science and Social Research, Kuala Lumpur, 2010.
[2] V. Malik, "ARIMA/GARCH (1,1) Modelling and Forecasting for A GE Stock Price Using R," ELK Asia Pacific Journal of Marketing and Retail Management, pp. 2349-2317, 2017.
[3] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," Journal of Econometrics, vol. 31, no. 3, pp. 307-327, 1986.
[4] M. Z. D. Shbier, "SWGARCH: An Enhanced Garch Model For Time Series Forecasting.,” Sintok, Kedah, 2017.
[5] X. Lu, D. Que and Q. Cao, "Volatility Forecast Based on the Hybrid Artificial Neural Network and GARCH-type Models," Procedia Computer Science, vol. 91, pp. 1044-1049, 2016.
[6] E. Hajizadeh, A. Seifi, M. Zarandi and I. Turksen, "A hybrid modeling approach for forecasting the volatility of S&P 500 index return," Expert Systems with Applications, vol. 39, no. 1, pp. 431-436, 2012.
[7] P. Kapoor and S. S. Bedi, "Weather Forecasting Using Sliding Window Algorithm," ISRN Signal Processing, pp. 1-5, 2013.
[8] E. Bingham, A. Gionis, N. Haiminen, H. Hiisilä, H. Mannila and E. Terzi, "Segmentation and dimensionality reduction," in SIAM International Conference on Data Mining, 2006.
[9] S. Hayashi, S. Ono, S. Hosoda, M. Numao and K.-i. Fukui, "Error Detection of Ocean Depth Series Data," in Machine Learning and Applications, Anaheim, CA, USA, 2016.
[10] T. R. R. M. Kapila, D. Šeneviratna, W. Jianguo and H. I. Arumawadu, "A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models," in In 2015 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC), 2015.