Multivariate generalized autoregressive score model (case study: vegetable oils and crude oil price data)

D Jayanti*, K Sadik, and F M Afendi
Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, West Java, INDONESIA
E-mail: dwi_9192@apps.ipb.ac.id

Abstract. A multivariate econometric model can be used to forecast volatilities and dynamic correlations between various assets. Volatilities and dynamic correlations forecasting are important applied in risk management, hedging, asset allocation, and asset pricing options. This study uses a multivariate Generalized Autoregressive Score (GAS) model to analyze and forecast volatilities and dynamic correlations of the weekly prices of Crude Palm Oil, coconut oil, soybean oil, and crude oil. The GAS model is a new framework based on a score-driven time series model for updating time-varying parameters. Another multivariate econometric model is the Dynamic Conditional Correlation (DCC) GARCH model as a comparison. The empirical experiments regarding the distribution of return data show that the data can be approximated by the multivariate t-student distribution. The likelihood ratio test in the multivariate GAS model shows that the time-varying parameters in the GAS model are volatility, correlation, and location parameters. The evaluation performance model based on RMSE and MAE show that the multivariate GAS model has better performance in estimating and forecasting volatilities than the DCC GARCH model. Multivariate GAS models have better performance in estimating and forecasting the dynamic correlation on returns of CPO and soybean oil.

1. Introduction
Vegetable oil is an oil extracted from plants. Vegetable oil is useful as edible oil for food (edible oil) and other than food (non-edible oil). Indonesia is a major producer of world vegetable oil, namely Crude Palm Oil (CPO) and coconut oil. The price of Indonesia's vegetable oil correlates with the other vegetable oil price, such as soybean oil. The interaction between vegetable oil prices is due to the use of substitutions between one type of vegetable oil and another [1]. Research by Arianto et al. in 2010, Alam et al. in 2019, and Priyati and Tyers in 2016 show that vegetable oil prices also correlate with crude oil prices [1] - [3]. The use of vegetable oil as raw material for biodiesel is one of the factors driving the link between vegetable oil prices and crude oil prices.

As the largest CPO supplier country in the world, Indonesia has a great opportunity as a price maker. However, the Indonesian government does not have the power to determine prices both in international and domestic markets [4]. Bursa Malaysia and Rotterdam in the Netherlands still control the reference price of CPO. The Indonesian price for coconut oil is also a reference to the Rotterdam price. Indonesia plays a role as a price taker in the international trade of CPO and coconut oil, thus causing the uncertainty of CPO and coconut oil prices in Indonesia and causes price volatility.

Supply and demand for vegetable oil and crude oil in the global market strongly influence the price. The existence of variations between supply and demand of vegetable oil and crude oil in the market over
time and unpredictable is called supply and demand shock. According to Gilbert and Morgan in 2010, supply and demand shock causes unpredictable prices and price volatility [5]. Price volatility encourages trading actors, including consumers, producers, exporters, importers, and speculators to do hedging of assets and risk management. They need a model of the volatility and the correlation between the prices of CPO, other vegetable oils, and crude oil to forecast them in the future. The forecast value of volatility and correlation is useful for risk management, hedging, asset allocation, and asset pricing options.

Classical time-series modeling is generally under the assumption of the constant variance of residual (homoscedasticity). However, time-series data have high volatility in particular periods and low in other periods so that the assumption of homogeneity of the variance is difficult to fulfill. The violation of heteroscedasticity becomes a problem in volatility modeling. Volatility and dynamic correlation modeling need to use a model to accommodate any violation of these assumptions. The classic model for volatility modeling is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. Stochastic volatility is another volatility modeling. Chan and Grant in 2016 used the SV and GARCH models to forecast oil price volatility [6]. Creal et al. in 2010 developed a model for modeling volatility, namely the generalized autoregressive score-driving model (GAS) [7]. However, most of the volatility modeling uses a single variable time series model. Univariate modeling lacks flexibility in explaining price volatility data. Therefore, this research is used multivariate volatility modeling.

The classical multivariate econometric model used for modeling volatility and dynamic correlation is the Dynamic Conditional Correlation (DCC) GARCH model. Pal and Mitra in 2019 used three types of multivariate GARCH models (DCC-GARCH, ADCC GARCH, and GO GARCH) to volatility and dynamic correlation modeling between crude oil price and agricultural commodity prices [8]. Ding and Vo in 2012 used the multivariate DCC-GARCH model and the Multivariate Stochastic Volatility (MSV) model to volatility modeling between oil price and exchange rates [9]. The result concluded that the multivariate DCC-GARCH model has better performance in forecasting volatility compared to the Multivariate Stochastic Volatility (MSV) model. Creal et al. in 2011 developed a data-driven model for time-varying parameters that is the Multivariate Generalized Autoregressive Score (GAS) model for time-varying volatilities and correlations [10]. The GAS model is a new framework based on a score-driven time series model for updating time-varying parameters. The GAS model has many advantages over classic volatility models. These advantages, for example, the likelihood can be evaluated directly, extensions to asymmetry, long memory. More complex dynamics can be considered without introducing further complexity. Chen and Xu in 2019 used a novel multivariate GAS model to model and forecast volatility and the correlation between oil and gold prices [11]. The results showed that the GAS multivariate model has good performance in volatility forecasting and dynamic correlation.

This study aims to predict and forecast the volatility and dynamic correlation between the prices of CPO, coconut oil, soybean oil, and crude oil. Modeling and forecasting are carried out by comparing the multivariate GAS and DCC-GARCH models.

2. Data Source
The data used in this study is secondary data obtained from the Indonesia Commodity & Derivatives Exchange (ICDX), the Commodity Futures Trading Supervisory Agency (Bappebti), and macrotrend.com. The price data used is the future closing price with a weekly period from the 1st week of 2010 to the 53rd week of 2019 for CPO, coconut oil, soybean oil, and crude oil (Brent). To validate the forecasting results, the data will be divided into training data (in of sample) and test data (out of sample). Training data is data from 2010 to 2017, while test data is data from 2018 to 2019. Modeling and forecasting volatility and dynamic correlation use the return data with the following formula

\[
 r_{i,t} = \ln \left( \frac{P_{i,t}}{P_{i,t-1}} \right) \times 100
\]

where:

- \( r_{i,t} \) = price returns of commodity-\( i \) at time-\( t \)
- \( P_{i,t} \) = weekly closing price of commodity-\( i \) at time-\( t \)
\[ P_{t-1} = \text{weekly closing price commodity-i at time-} t-1 \]
\[ i = 1, 2, 3, \text{and 4 refer to CPO, coconut oil, soybean oil, and crude oil, respectively.} \]

3. Methodology

The multivariate GAS model

The generalized autoregressive score (GAS) model is an observation-driven model based on the score function of the predictive model density at time-\( t \). The basic idea of the multivariate GAS model is to update the time-varying parameters based on the information available in time-\( t \). Let the return data is a \( N \)-dimensional vector at time \( t \) with conditional distribution as follows:

\[ r_t | r_{1:t-1} \sim p(r_t; f_t, \theta) \] (1)

\[ r_{1:t-1} \equiv (r_t', r_{t-1}', \ldots, r_1')' \] is contains the past value of \( r_t \) up to time t-1. \( f_t \) is the vector time-varying parameter and \( \theta \) is a static parameter vector. The time-varying parameter vector \( f_t \) are driven by the scaled score function of the conditional observation density in equation (1).

For simplicity, this study uses a first-order multivariate GAS model with \( p = 1 \) and \( q = 1 \) or multivariate GAS (1,1) model. Therefore, the updating mechanism of the time-varying parameter vector \( f_t \) is given by autoregressive updating equation follows as:

\[ f_{t+1} \equiv \kappa + A(\theta) s_t + B(\theta) f_t \] (2)

where:
\( \kappa, A(\theta) \) and \( B(\theta) \) are matrices of coefficients with proper dimensions, and \( s_t \) is scaled score function derivative of the density function (1) at the time \( t \) with respect to the time-varying parameter vector \( f_t \), as follows:

\[ s_t \equiv S_t \nabla_t \]
\[ \nabla_t \equiv \frac{\partial \ln p(r_t | f_{t-1}; f_t, \theta)}{\partial f_t} \] (3)

\( S_t \) is a positive definite scaling matrix known at time \( t \). Via the choice of scaling matrix \( S_t \), the GAS model allows more flexibility in the score function is used for updating \( f_t \) [12]. \( \nabla_t \) is the score of (1) evaluated at \( f_t \). Scaling matrix that make the GAS model encompasses the well-known classic volatility model such us GARCH as follows:

\[ S_t \equiv J_{t|t-1}^{-1} \text{dan } J_{t|t-1} \equiv E_{t-1}[\nabla_t \nabla_t'] \]

where \( J_{t|t-1} \) is information matrix of \( f_t \). Parameter \( \gamma \) usually takes value in the set \( \{0, \frac{1}{2}, 1\} \) and fixed by the researcher. Creal et al. (2013) mention that the quantity of \( s_t \) indicated the direction to update the vector of parameter \( f_t \) to \( f_{t+1} \), acting as a steepest ascent algorithm for improving the model local fit given the current parameter position. This updating procedure resembles the Newton-Raphson algorithm.

The parameter vector \( f_t \) in (2) has linear specification and not restricted. In practical, the time-varying parameter \( f_t \) such as variance, correlation and degree of freedom number is often restricted. The solution of this problem under the GAS model is to use a non-linear link function that map \( f_t \) into \( f_t \) and \( f_t \) has a linear dynamic specification of (2) [13].

The GAS model has a useful property that the time-varying parameter vector \( f_t \) is perfectly predictable and the log-likelihood function can be easily evaluated with the prediction error decomposition. In other words, the time-varying parameter vector \( f_t \) can be estimated by the Maximum Likelihood Estimation (MLE) method [14].

The multivariate DCC-GARCH model

The DCC GARCH model was developed by Engle in 2002 with the basic idea is decomposed the covariance matrix into a conditional standard deviation matrix and a conditional correlation matrix [15]. The DCC-GARCH models are defined as:

\[ r_t = \mu + a_t \] (4)
\[ a_t = H_t^{1/2} z_t \]  
\[ H_t = D_t R_t D_t \]  

Where:
- \( r_t \) = n x 1 return data vector of n assets at time t
- \( a_t \) = n x 1 mean-corrected return vector of n assets at time t with \( E[a_t]=0 \) and \( \text{Cov}[a_t]=H_t \)
- \( \mu_t \) = n x 1 the expected value vector of the conditional \( r_t \)
- \( H_t \) = n x n conditional covariance matrix of \( a_t \) at time t
- \( D_t \) = n x n conditional standard deviation matrix of \( a_t \) at time t
- \( R_t \) = n x n conditional correlation matrix of \( a_t \) at time t
- \( z_t \) = n x 1 vector of iid errors such that \( E[z_t]=0 \) and \( \text{Cov}[z_t]=I \)
- \( \mu_t \) in mean model (4) can be a constant or a time series model such as AR (p), MA (q), or ARMA (p,q).
  In this study, the average model used is the time series model from the identification results.

The element of the diagonal matrix of conditional standard deviation \( D_t \) is standard deviations from a univariate variance model such as the GARCH model.

\[ D_t = \begin{bmatrix} \sqrt{h_{1,t}} & 0 & \ldots & 0 \\ 0 & \sqrt{h_{2,t}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{h_{n,t}} \end{bmatrix} \]

For simplicity, this research uses the univariate GARCH (1,1) model specification, such as:
\[ h_{i,t} = \alpha_0 + \alpha a_{i,t-1}^2 + \beta h_{i,t-1} \]

\( R_t \) is a symmetric time varying correlation matrix.

\[ R_t = \begin{bmatrix} 1 & \rho_{12,t} & \ldots & \rho_{1n,t} \\ \rho_{12,t} & 1 & \ldots & \rho_{2n,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n,t} & \rho_{2n,t} & \ldots & 1 \end{bmatrix} \]

The requirements have to be considered when specifying a form of \( R_t \):
- Matrix \( R_t \) has to be definite positive to ensure Matrix \( H_t \) has to be a definite positive because it is a covariance matrix.
- All element of matrix \( R_t \) has to be equal or less than one by definition.

To ensure the requirement of the DCC Model, \( R_t \) is decomposed into:
\[ R_t = Q_t^{-1} \tilde{Q}_t Q_t^{-1} \]
with \( \tilde{Q}_t = (1-a-b)Q_t + \alpha \tilde{e}_{t-1} \tilde{e}_{t-1}' + bQ_{t-1} \) 

The method used in estimating the conditional variance matrix is the two-stage Quasi Maximum Likelihood Method [16]. The first step is to estimate the univariate GARCH parameters and the second stage is to estimate the parameters of the correlation structure model.

**Evaluation Forecasting Performance**
Forecasting performance is evaluated using the smallest RMSE and MAE of the two models. The RMSE and MAE formulas are following Poon (2005) and Chen and Xu (2019) as follow [17] - [18]:

\[ \text{MAE}_\sigma = \frac{1}{N} \sum_{t=1}^{N} |\hat{\sigma}_t - \sigma_t| \]
\[ \text{MAE}_\rho = \frac{1}{N} \sum_{t=1}^{N} |\hat{\rho}_t - \rho_t| \]
\[ \text{RMSE}_\sigma = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (\hat{\sigma}_t - \sigma_t)^2} \]
\[ \text{RMSE}_\rho = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (\hat{\rho}_t - \rho_t)^2} \]
\( \hat{\sigma}_t \) and \( \hat{\rho}_t \) is 1-step ahead rolling forecast results of volatility and correlation at period \( t \) for each model. \( \sigma_t \) and \( \rho_t \) is volatility and correlation realization as proxies of the actual volatility and correlation. Volatility is a latent variable. Therefore, in the evaluation of volatility forecasting, several measures are often used as a proxy for the actual volatility. This study uses the absolute value of return as a proxy for actual volatility because the absolute value of returns can predict volatility well [18]. Mapa et al. used a rolling correlation with specific periods as a proxy for the actual correlation [19]. Therefore, this study using a rolling correlation of daily price returns of two commodity price returns with a period of 5 weeks as actual correlation.

4. Empirical result

The data exploration and preliminary analysis

The price returns for the four commodities fluctuated throughout the study period. The returns of CPO and coconut oil are of extreme value at several points in time. In 2015, the price of CPO decreased drastically, resulting in high return value and a negative sign. The decline in CPO prices during this period was caused by reduced demand from exporting countries, an oversupply of soybean oil that is a substitute commodity, the ineffective biodiesel program in Indonesia and Malaysia, and the decline in crude oil prices.

The returns range differs between periods. The difference returns range is seen in the return of crude oil prices (Figure 1). In 2010-2014, the returns range of crude oil was -8 to 7. In the following period, the returns range was between -15 to 15. The difference in the returns range is an initial indication of the non-constant variance of error.

The scatter plot between the returns from the four commodities along with the Pearson correlation is presented in Figure 2. Based on this figure, there is a significant relationship between the six pairs of returns. The strongest correlation is between CPO and soybean oil, which is 0.5. The high correlation between them is due to the use of the two commodities which are mutually substituting. The weakest correlation is between the returns of palm oil and crude oil.

Early identification of dynamic correlation can be done using a rolling correlation throughout the study period. A 15-week rolling correlation shows a fluctuating correlation that tends to be positive. The 15-week rolling correlation between returns of CPO and soybean oil tends to be more consistently positive than other pairs. The existence of significant correlations justifies the reasons for conducting analyzes of volatility and correlation simultaneously.
Figure 1. Time series plots of returns for CPO, coconut oil, soybean oil, and crude oil
Table 1 shows descriptive statistics of the price returns for the four commodities. The standard deviation is a measure of unconditional volatility. The largest standard deviation is coconut oil return of 4.13, while the other three returns range from 2 to 3.5. That indicates that the return of coconut oil has the highest volatility compared to other commodities.

The skewness of CPO and crude oil is negative, while the skewness of coconut oil and crude oil is positive. The kurtosis value of the return of the four commodities is less than 3, which means that the
data distribution is platikurtic. It value indicates that the distribution of returns on CPO and crude oil is left-skewed and platikurtic, while the distribution of returns for soybean oil and coconut oil is right-skewed and platikurtic.

**Table 1.** Descriptive statistics for returns of CPO, coconut oil, soybean oil, and crude oil

| Descriptive Statistics | CPO  | Coconut Oil | Soybean Oil | Crude Oil |
|------------------------|------|-------------|-------------|-----------|
| Mean                   | 0.08 | 0.06        | -0.03       | -0.04     |
| Median                 | 0.13 | 0.03        | -0.14       | 0.11      |
| Standard Deviation     | 2.51 | 4.13        | 2.22        | 3.30      |
| Minimum                | -10.62 | -16.09 | -6.16       | -13.73    |
| Maximum                | 13.11 | 20.69       | 7.19        | 12.42     |
| Skewness               | -0.02 | 0.23        | 0.19        | -0.09     |
| Kurtosis               | 2.64 | 2.55        | 0.19        | 1.50      |
| Observations           | 521.00 | 521.00 | 521.00      | 521.00    |

The univariate normality test on returns using the Jarque-Bera test rejects the null hypothesis on returns of CPO, coconut oil, and crude oil. It means that the three returns do not follow the normal distribution, while returns of soybean oil follow the normal distribution at the 1% significant level. The stationarity test using the Dickey-Fuller test shows that the four returns series are stationary at the 1% significant level. However, all four returns series exists an autocorrelation based on the Ljung-box test at lag 20. The ARCH test on error returns from the mean rejects the null hypothesis that there is no ARCH effect. It shows that the error has a non-constant variance. The results of these tests confirm the use of the multivariate GAS model and the DCC-GARCH model for forecasting the volatility of all returns series.

![Empirical simulation result for multivariate normality test](image)

The multivariate normality test on returns using the Doornik Hansen test shows the rejection of the null hypothesis that the distributions of returns are symmetric [20]. An empirical experiment will be
conducted to find the distribution of the data return approach. The basic idea of this empirical illustration is the distribution of the sample approachable with the distribution of the majority of the subsample [11].

First, we divide the data into sub-samples with a fixed size and sequential in chronological order. This experiment divided the sample into sub-samples with sizes from 20 to 400 with the addition of 10. Second, we perform the Doornik-Hansen test for each sub-sample. Third, we calculate the rejection rate from the test results. The rejection rate is the ratio of the number of sub-samples that reject the hypothesis to the total number of sub-samples at a fixed sub-sample size. The high rejection rate means that majority of the sub-samples did not follow the multivariate normal distribution.

The results show that the smaller the sub-sample size, the smaller the rejection rate (Figure 3). It means that the smaller the sample size, the more subsamples that follow the multivariate normal distribution. Based on the results of these experiments, the distribution of the returns can be approximated by the multivariate student's t distribution. Therefore the parameter estimation of the multivariate GAS model and the DCC-GARCH model is based on the multivariate student's t distribution error.

The estimation result of multivariate GAS model
The multivariate GAS model is a data-driven model for modelling time-varying parameters. The mechanism updating of the time-varying parameter is using the autoregressive function. There are four types of time-varying parameters for the multivariate GAS model based on the multivariate student's t distribution error (the multivariate t-GAS model), namely location ($\mu$), volatility ($\phi$), correlation ($\rho$), and shape ($\nu$). The selection of time-varying parameters in the model from the four types of parameters in the multivariate t-GAS used the likelihood ratio test. This study performs three stages of the likelihood ratio test for testing the correlation, location, and shape as the time-varying parameter. The basic multivariate GAS model is a model with volatility as a time-varying parameter.

| No | Testing Purpose                      | Hypothesis             | Chi square value | P-value |
|----|-------------------------------------|------------------------|------------------|---------|
| 1  | Testing correlation as a time-varying parameter | H0: Model 1            | 5.8292           | 0.0542  |
|    |                                      | H1: Model 2            |                  |         |
| 2  | Testing shape as a time-varying parameter   | H0: Model 2            | 0.0020           | 0.9990  |
|    |                                      | H1: Model 3            |                  |         |
| 3  | Testing location as a time-varying parameter  | H0: Model 3            | 29.6075          | 0.0000  |
|    |                                      | H1: Model 4            |                  |         |

Table 2. The likelihood ratio test for selection of the best multivariate t-GAS model

Note:
Model 1: t-GAS model with volatility as time-varying parameters
Model 2: t-GAS model with volatility and correlation as time-varying parameters
Model 3: t-GAS model with volatility, correlation, and shape as time-varying parameters
Model 4: t-GAS model with volatility, correlation, shape, and locations as time-varying parameters

The testing of correlation parameters as time-varying parameters results in rejection of the null hypothesis at the 10% significant level (Table 2). Multivariate t-GAS models with volatility and correlation as time-varying parameters are better than models with only volatility as time-varying parameters. The location parameter is also significant as a time-varying parameter with a p-value that is close to zero. Meanwhile, the shape parameter is not statistically significant as the time-varying parameter. Based on the likelihood ratio test, the best multivariate t-GAS model is a model with volatility, correlation, and location as time-varying parameters.

The coefficients of the autoregressive model for updating the time-varying parameters are $\kappa$, $A$, and $B$. The matrix $\kappa$ is a diagonal matrix that contains constants that correspond to time-varying parameters.
of volatility (scale), correlation, shape, and location and static parameters. The diagonal elements of the matrix $\kappa$ are as follows:

\[
\begin{pmatrix}
\kappa_{\mu 1}, \kappa_{\mu 2}, \kappa_{\mu 3}, \kappa_{\phi 1}, \kappa_{\phi 2}, \kappa_{\phi 3}, \kappa_{\rho 12}, \kappa_{\rho 13}, \kappa_{\rho 24}, \kappa_{\rho 24}, \kappa_{\rho 34}, \kappa_{\nu}
\end{pmatrix}
\]

Matrix $A$ is a diagonal matrix with diagonal elements containing the coefficients of the scale score vectors in the previous period $s_{t-1}$. Matrix $B$ is a diagonal matrix with a diagonal element containing the coefficients of the time-varying parameter vectors in the previous period $f_{t-1}$.

\[
A = \text{diag} \left( a_{\mu 1}, a_{\mu 2}, a_{\mu 3}, a_{\phi 1}, a_{\phi 2}, a_{\phi 3}, a_{\rho 12}, a_{\rho 13}, a_{\rho 24}, a_{\rho 24}, a_{\rho 34}, 0 \right).
\]

\[
B = \text{diag} \left( b_{\mu 1}, b_{\mu 2}, b_{\mu 3}, b_{\phi 1}, b_{\phi 2}, b_{\phi 3}, b_{\rho 12}, b_{\rho 13}, b_{\rho 24}, b_{\rho 24}, b_{\rho 34}, 0 \right).
\]

With the reparameterization process using the non-linear link function on vector $f$, matrices $A$ and $B$ become as follows:

\[
A = \text{diag} \left( a_{\mu}, a_{\mu}, a_{\mu}, a_{\mu}, a_{\phi}, a_{\phi}, a_{\phi}, a_{\phi}, a_{\rho}, a_{\rho}, a_{\rho}, a_{\rho}, 0 \right)
\]

### Table 3. Estimation results of selected multivariate t-GAS (1,1) model

| Parameters | Estimate | Std. Error | t-value | p-value |
|------------|----------|------------|---------|---------|
| $\kappa_{\mu 1}$ | 0.1016 | 0.1165 | 0.8726 | 0.1914 |
| $\kappa_{\mu 2}$ | -0.0487 | 0.1154 | -0.4224 | 0.3364 |
| $\kappa_{\mu 3}$ | 0.1008 | 0.1330 | 0.7576 | 0.2244 |
| $\kappa_{\mu 4}$ | 0.2280 | 0.1690 | 1.3495 | 0.0886 |
| $\kappa_{\phi 1}$ | 0.0445 | 0.0153 | 2.9108 | 0.0018 |
| $\kappa_{\phi 2}$ | 0.0500 | 0.0161 | 3.0969 | 0.0010 |
| $\kappa_{\phi 3}$ | 0.0551 | 0.0187 | 2.9418 | 0.0016 |
| $\kappa_{\phi 4}$ | 0.0708 | 0.0231 | 3.0696 | 0.0011 |
| $\kappa_{\rho 12}$ | 0.3609 | 0.3062 | 1.1788 | 0.1192 |
| $\kappa_{\rho 13}$ | 0.4516 | 0.3838 | 1.1759 | 0.1198 |
| $\kappa_{\rho 14}$ | 0.3541 | 0.3009 | 1.1769 | 0.1196 |
| $\kappa_{\rho 23}$ | 0.4052 | 0.3439 | 1.1782 | 0.1194 |
| $\kappa_{\rho 24}$ | 0.4571 | 0.3848 | 1.1877 | 0.1175 |
| $\kappa_{\rho 34}$ | 0.5219 | 0.4424 | 1.1797 | 0.1191 |
| $\kappa_{\nu}$ | -1.5570 | 0.3721 | -4.1846 | 0.0000 |
| $a_{\mu}$ | 0.5474 | 0.1115 | 4.9094 | 0.0000 |
| $a_{\phi}$ | 0.0455 | 0.0074 | 6.1560 | 0.0000 |
| $a_{\rho}$ | 0.0179 | 0.0122 | 1.4632 | 0.0717 |
| $b_{\mu}$ | 0.0559 | 0.1570 | 0.3560 | 0.3609 |
| $b_{\phi}$ | 0.9415 | 0.0185 | 50.9374 | 0.0000 |
| $b_{\rho}$ | 0.6645 | 0.2842 | 2.3383 | 0.0097 |

Note:
Subscript 1, 2, 3, and 4 refer to CPO, coconut oil, soybean oil, and crude oil, respectively.

The estimation result of the multivariate t-GAS model is presented in the Table 3. The coefficient $b_{\phi}, b_{\rho},$ and $a_{\phi}$ are statistically significant at the 1% level, while $a_{\rho}$ significant at a higher level, namely 10%. The multivariate t-GAS model can capture the high volatility of returns represented by a
coefficient $b_\phi$ close to 1 and a high level of significance. However, the multivariate t-GAS model cannot capture a strong and consistent correlation between returns. It is supported by a small value of $a_\rho$ and a lower level of significance.

The estimation result of multivariate DCC-GARCH model
Estimating parameters of the DCC GARCH model (1,1) was carried out in two stages. The first stage estimates the variance model parameters with univariate GARCH on each return series. The second stage estimates the correlation model on the DCC GARCH (1,1). The mean model needs to be identified precisely so that the variance model accurately models the fluctuation of returns to the average model.

| Commodity   | The Tentative Model | AIC       | The Best Model   |
|-------------|---------------------|-----------|------------------|
| CPO         | ARIMA(0,0,2)        | 1903.84   | ARIMA(1,0,0)     |
|             | ARIMA(1,0,0)        | 1901.95   |                   |
| Coconut Oil | ARIMA(0,0,1)        | 2365.38   | ARIMA(0,0,1)     |
|             | ARIMA(2,0,0)        | 2365.84   |                   |
| Soybean Oil | ARIMA(0,0,1)        | 1860.55   | ARIMA(0,0,1)     |
|             | ARIMA(2,0,0)        | 1860.68   |                   |
| Crude Oil   | ARIMA(0,0,1)        | 2140.56   | ARIMA(0,0,1)     |
|             | ARIMA(1,0,0)        | 2143.71   |                   |

The model means were identified using Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), Extended Autocorrelation Function (EACF), and auto ARIMA in the R program. The results are presented in Table 4. The best mean model is selected based on the smallest AIC from several tentative models. The best mean model for returns of CPO is the AR (1) model, while the other returns are the MA (1) model.

| Commodity   | Variance Model Specifications | Criteria |
|-------------|-------------------------------|----------|
|             |                               | Akaike   | Bayes   | Shibata | Hannan-Quinn |
| CPO         | GARCH (1,1)                   | 4.5487   | 4.5972  | 4.5485  | 4.5679       |
|             | eGARCH (1,1)                  | 4.5369   | 4.5950  | 4.5365  | 4.5599       |
| Coconut Oil | GARCH (1,1)                   | 5.5226   | 5.5711  | 5.5223  | 5.5418       |
|             | eGARCH (1,1)                  | *        | *       | *       | *            |
| Soybean Oil | GARCH (1,1)                   | 4.4789   | 4.5274  | 4.4787  | 4.4981       |
|             | eGARCH (1,1)                  | 4.4818   | 4.5400  | 4.4814  | 4.5048       |
| Crude Oil   | GARCH (1,1)                   | 4.9440   | 4.9925  | 4.9438  | 4.9632       |
|             | eGARCH (1,1)                  | 4.9083   | 4.9665  | 4.9079  | 4.9313       |

Besides the standard GARCH (1,1), the variance model considers the asymmetric GARCH model such as e-GARCH (1,1). The consideration of using the e-GARCH (1,1) model is the return data that is often asymmetrical. The selection of variance models between the standard GARCH (1,1) and e-GARCH (1,1) uses four criteria, namely Akaike, Bayes, Shibata, and Hannan-Quinn. The values of the four criteria for the two variance models are presented in table 5. The appropriate variance model for
returns of CPO and crude oil is the e-GARCH (1,1) model, while the variance model for the other two returns is the standard GARCH (1,1).

Table 6. Estimation result of selected DCC GARCH (1,1) model

| Parameter | Estimate | Std. error | t-value | p-value |
|-----------|----------|------------|---------|---------|
| $\mu_1$  | 0.0150   | 0.1678     | 0.0894  | 0.9288  |
| $\phi_{11}$ | 0.3287   | 0.0298     | 11.0260 | 0.0000  |
| $\omega_1$ | 0.2900   | 0.1006     | 2.8840  | 0.0039  |
| $\alpha_{11}$ | -0.1147  | 0.0571     | -2.0063 | 0.0448  |
| $\beta_{11}$ | 0.8312   | 0.0581     | 14.3094 | 0.0000  |
| $\gamma_1$ | 0.2221   | 0.0767     | 2.8962  | 0.0038  |
| $\mu_2$  | 0.2101   | 0.2241     | 0.9375  | 0.3485  |
| $\theta_{21}$ | 0.2966   | 0.0468     | 6.3414  | 0.0000  |
| $\omega_2$ | 1.6485   | 0.7185     | 2.2943  | 0.0218  |
| $\alpha_{21}$ | 0.2636   | 0.0951     | 2.7723  | 0.0056  |
| $\beta_{21}$ | 0.6595   | 0.0850     | 7.7611  | 0.0000  |
| $\mu_3$  | -0.0732  | 0.1392     | -0.5260 | 0.5989  |
| $\theta_{31}$ | 0.2576   | 0.0538     | 4.7850  | 0.0000  |
| $\omega_3$ | 0.8827   | 0.8057     | 1.0956  | 0.2733  |
| $\alpha_{31}$ | 0.0553   | 0.0391     | 1.4143  | 0.1573  |
| $\beta_{31}$ | 0.7691   | 0.1750     | 4.3938  | 0.0000  |
| $\mu_4$  | -0.1469  | 0.1689     | -0.8700 | 0.3843  |
| $\theta_{41}$ | 0.3489   | 0.0487     | 7.1694  | 0.0000  |
| $\omega_4$ | 0.0401   | 0.0090     | 4.4813  | 0.0000  |
| $\alpha_{41}$ | -0.1064  | 0.0263     | -4.0376 | 0.0001  |
| $\beta_{41}$ | 0.9807   | 0.0015     | 664.0501| 0.0000  |
| $\gamma_4$ | 0.1939   | 0.0603     | 3.2162  | 0.0013  |
| $\alpha_4$ | 0.0341   | 0.0169     | 2.0122  | 0.0442  |
| $b_1$    | 0.6403   | 0.1290     | 4.9630  | 0.0000  |
| $\nu$    | 14.8049  | 3.9454     | 3.7524  | 0.0002  |

Note:
Subscript 1, 2, 3, and 4 refer to CPO, coconut oil, soybean oil, and crude oil, respectively.

All coefficients are significant at the 5% level except for the constant on the mean model of each commodity and the coefficient $\alpha_{31}$ on the variance model for returns of soybean oil. It shows that returns of soybean oil have a short term memory while other returns have a long term memory. All coefficients $\beta$ on the variance model are significant at the 1% level so that this model can capture the existence of volatility.

The estimator for $\alpha_{4}$ and $b_1$ in the correlation structure equation is 0.0341 and 0.6403, respectively, which are significant at the 5% level. It shows a statistically significant correlation between returns. The existence of a dynamic correlation between the price returns of these four commodities is not strong enough because the sum of $\alpha_4 + b_1$ is 0.6743 and is not close to 1.
The fit performance on in-of-sample data and the forecasting performance on out-of-sample data

The evaluation of the best model for estimating and forecasting volatility and the dynamic correlation between the multivariate t-GAS and t-DCC GARCH models use the RMSE and MAE values. The best model has the smallest RMSE and MAE values. The RMSE and MAE values in estimating volatility and the correlation in the in-of-samples data are presented in Table 7. Evaluation for forecasting volatility and correlation in the out-of-sample data are presented in Table 8.

Table 7. The comparison of RMSE and MAE for estimating and forecasting volatility on in-of-sample data and out-of-sample data

| Volatility | Multivariate t-GAS (1,1) Model | DCC GARCH (1,1) Model |
|------------|--------------------------------|-----------------------|
|            | RMSE | MAE | RMSE | MAE |
| In-of-sample |      |     |      |     |
| CPO        | 1.5928 | 1.2842 | 1.7131 | 1.3626 |
| Coconut Oil | 2.5966 | 2.0735 | 3.0012 | 2.3277 |
| Soybean Oil | 1.4019 | 1.1958 | 1.4301 | 1.1895 |
| Crude Oil  | 1.9730 | 1.6018 | 2.2305 | 1.7583 |
| Out-of-sample |      |     |      |     |
| CPO        | 1.7788 | 1.2599 | 1.8300 | 1.3426 |
| Coconut Oil | 2.6552 | 2.0200 | 2.8841 | 2.1914 |
| Soybean Oil | 1.1563 | 0.9650 | 1.2697 | 1.0895 |
| Crude Oil  | 2.1443 | 1.6970 | 2.1882 | 1.7831 |

The multivariate t-GAS model produces RMSE and MAE values that are smaller than the GARCH t-DCC model in estimating and forecasting volatility in the four returns. Based on these results, the multivariate t-GAS model has better performance in estimation and forecasting volatility. It is in line with the results of the research by Chen and Xu in 2019 that stated that the multivariate t-GAS model has good performance in forecasting volatility [11].

The estimate and forecast values of soybean oil return volatility produce the smallest RMSE and MAE. The largest RMSE and MAE occur in estimating and forecasting the volatility of coconut oil. It shows that the returns of coconut oil have a complex movement and make the error of the two models in forecasting is still large. The value of the volatility forecast for the four returns series based on the multivariate t-GAS model tends to be lower than the t-DCC GARCH model. Similar results also occur in the volatility estimation of the in-of-sample data. However, in general, the forecast results for the two models are quite close. The difference in forecast results between the two models is high when the return is far from the other values. The t-DCC GARCH model tends to be reactive to extreme returns. The forecast results based on this model are not robust on extreme values. It is consistent with the statement of Creal et al. in 2012 that the multivariate t-GAS model produces robust forecast values [12]. The comparison of the results of estimating volatility between the multivariate t-GAS and t-DCC GARCH models is presented in Figure 3, while the comparison of the volatility forecasting results is presented in Figure 4.
Figure 4. Multivariate t-GAS model and t-DCC GARCH volatility estimate result comparison
Figure 5. Multivariate t-GAS model and t-DCC GARCH volatility forecast result comparison
The pair of CPO and soybean oil returns series have the strongest dynamic correlation compared to other pairs. The dynamic correlation between the returns of the two commodities ranges from 0.4 to 0.6 throughout the study period. The estimate and forecast values of the dynamic correlation between the two returns based on the t-DCC GARCH model tend to be higher than the multivariate t-GAS model. However, in the other pairs, the estimate and forecast values based on the multivariate t-GAS model were higher than the t-DCC GARCH model.

The evaluation of the best model for estimating and forecasting the correlation between returns is carried out based on the RMSE and MAE values (Table 8). The multivariate t-GAS model has better performance than the t-DCC GARCH model in estimating and forecasting the correlation on the pair of CPO and soybean oil returns series. However, in the other five pairs, the GARCH t-DCC model produced smaller RMSE and MAE in both estimating and forecasting the correlation.

Table 8. The comparison of RMSE and MAE for estimating and forecasting the correlation on in-of-sample data and out-of-sample data

| Correlation                     | Multivariate t-GAS (1,1) Model | DCC GARCH (1,1) Model |
|---------------------------------|---------------------------------|-----------------------|
|                                 | RMSE   | MAE   | RMSE   | MAE   |
| **In-of-sample**                |        |       |        |       |
| CPO and Coconut Oil             | 0.3743 | 0.3062| **0.3491** | **0.2824** |
| CPO and Soybean Oil             | **0.4694** | **0.4223** | 0.5051 | 0.4603 |
| CPO and Crude Oil               | 0.2882 | 0.2371| **0.2466** | **0.1987** |
| Coconut Oil and Soybean Oil     | 0.4540 | 0.4082| **0.4221** | **0.3758** |
| Coconut Oil and Crude Oil       | 0.3122 | 0.2542| **0.2738** | **0.2187** |
| Soybean Oil and Crude Oil       | 0.2512 | 0.1980| **0.2331** | **0.1862** |
| **Out-of-sample**               |        |       |        |       |
| CPO and Coconut Oil             | 0.3959 | 0.3491| **0.3836** | **0.3424** |
| CPO and Soybean Oil             | **0.4271** | **0.3780** | 0.4804 | 0.4320 |
| CPO and Crude Oil               | 0.2677 | 0.2100| **0.2347** | **0.1832** |
| Coconut Oil and Soybean Oil     | 0.4614 | 0.4185| **0.4227** | **0.3768** |
| Coconut Oil and Crude Oil       | 0.2519 | 0.2087| **0.2426** | **0.1953** |
| Soybean Oil and Crude Oil       | 0.2715 | 0.2251| **0.2663** | **0.2185** |

The performance in correlation forecasting of the multivariate t-GAS and t-DCC GARCH models is influenced by the consistency of correlation and data structure that can be reflected in the correlation structure coefficient in the model [10]. The correlation structure coefficient $\alpha_1$ in the t-DCC GARCH model is 0.0341. This value is greater than the coefficient of the correlation structure $\alpha_p$ in the multivariate t-GAS model that is 0.0179. The coefficient $\alpha_1$ in the t-DCC GARCH model has a higher significance with a p-value of 0.0442, while the coefficient $\alpha_p$ on the multivariate t-GAS model has a p-value of 0.0717. It shows that t-DCC GARCH model is more responsive in detecting the presence of dynamic correlation on price return data.

5. Conclusions

The returns of coconut oil have the highest estimated volatility compared to other return series. These returns have a complex movement characterized by the largest RMSE and MAE values compared to other returns. The strongest estimated dynamic correlation is the correlation between the returns of CPO and soybean oil. The correlation between the two returns ranges from 0.4 to 0.6.
The multivariate t-GAS model produces RMSE and MAE values that are smaller than the GARCH t-DCC model in estimating and forecasting volatility in the four returns. The multivariate t-GAS model has better performance than the t-DCC GARCH model in estimating and forecasting volatility in the four returns series. However, the multivariate t-GAS model performed well only in estimating and forecasting the correlation between returns for CPO and soybean oil. The rolling correlation between CPO and soybean oil tends to be more consistent than the other pairs supporting these results.

Modelling of volatility and dynamic correlation to hedge assets, risk management, and asset allocation is better to use a multivariate GAS model. This model is more robust in extreme returns, is more practical in its application, and has a good performance in volatility forecasting. For further research, it is better to include exogenous variables and asymmetric elements in the GAS model, considering the increasingly complex data structure.

References
[1] Arianto M E, Daryanto A, Arifin B and Nuryartono N 2010 Analisis Harga Minyak Sawit, Tinjauan Kointegrasi Harga Minyak Nabati Dan Minyak Bumi. J Manaj Agribisnis. 7(1)1–15 doi:10.17358/jma.7.1.1-15.
[2] Alam IA, Hairani H, and Singagerda FS 2019 Price Determination Model of World Vegetable and Petroleum. Int J Energy Econ Policy. 9(5) 157–177.
[3] Priyati RY 2016 Price Relationships in Vegetable Oil and Energy Markets. Annual Australasian Development Economics Workshop. Discussion Paper 16.11.
[4] Hafizah D 2011 Kajian Kebijakan Pemerintah Indonesia Dalam Perdagangan Cpo Indonesia Menggunakan Pendekatan Analisis Integrasi Pasar. J AGRISEP. 10(2) 154–170. doi:10.31186/jagrisep.10.2.154-170.
[5] Gilbert C and Morgan W 2010 Has food price volatility risen? Department of Economics, University of Trento, Italia. Report No.: 1002.
[6] Chan J and Grant A 2016 Modeling energy price dynamics: GARCH versus stochastic volatility. Energy Econ. 54(C):182–189.
[7] Creal D D, Koopman S J and Lucas A 2010 “Generalized Autoregressive Score Models With Applications”, working paper, University of Chicago Booth School of Business, pp 553-555.
[8] Pal D and Mitra S K 2019 Correlation dynamics of crude oil with agricultural commodities: A comparison between energy and food crops, Economic Modelling, 82 453-466.
[9] Ding L and Vo M 2012 Exchange rates and oil prices: A multivariate stochastic volatility analysis, The Quarterly Review of Economics and Finance, 52(1) 15-37.
[10] Creal D, Koopman S J and Lucas A 2011 A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, Journal of Business and Economic Statistics, 29(4) 552-563.
[11] Chen R and Xu J 2019 Forecasting volatility and correlation between oil and gold prices using a novel multivariate GAS model, Energy Economics, 78 379-391.
[12] Creal DD, Koopman S J, and Lucas A 2012. Generalized autoregressive score models with applications. Journal of Applied Econometrics. 28 777-795.
[13] Ardia D, Boudt K, and Catania L. 2019 Generalized Autoregressive Score Models in R: The GAS Package, Journal of Statistical Software, 88(1) 1-28.
[14] Blasques F, Koopman S J and Lucas A 2014 Maximum Likelihood Estimation for Generalized Autoregressive Score Models, Tinbergen Institute Discussion Paper, No. 14-029/III, Tinbergen Institute, Amsterdam and Rotterdam.
[15] Engle R F 2002 Dynamic conditional correlation – A simple class of multivariate GARCH models, Journal of Business and Economic Statistics, 20(3) 339-350.
[16] Orskaug E 2009 Mutlivariate DCC-GARCH Model With Variaous Error Distributions. [Thesis]. Norwegian University of Science and Technology.
[17] Poon SH 2005 A practical guide to forecasting financial market volatility. Chichester: Jhon Wiley and Sons, Ltd.
[18] Forsberg L, Ghysels E 2007 Why Do Absolute Returns Predict Volatility So Well? Journal of Financial Econometrics 5 (1) 31-67.
[19] Mapa D S, Paz N J I, Eustatquio J D and Mindanao A C 2014 Forecasting TimeVarying Correlation using the Dynamic Conditional Correlation (DCC) Model. Technical Report Paper MPRA No. 55861.
[20] Doornik J, Hansen H 2008 An omnibus test for univariate and multivariate normality. Oxford Bulletin of Economics and Statistics. 70 915-925.