Available online at
https://journal.uitm.edu.my/ojs/index.php/JEEIR

e-ISSN: 2289-2559

Journal of Emerging Economies & Islamic Research 12(2) 2024, 2787.

Journal of Emerging Economies and Islamic Research

Synergizing ARDL and LSTM methods for enhanced crude palm oil price forecasting in Malaysia

Mohd Shahrin Bahar[a*], Imbarine Bujang[b] , Abdul Aziz Karia[c], Nurzahidah Baharudin[d]

[a,d]*Faculty of Business and Management, Universiti Teknologi MARA*
[b,c]*Faculty of Business and Management, Universiti Teknologi MARA (UiTM Sabah)*

ARTICLE INFO

Article history:
Received 5 January 2024
Accepted 25 April 2024
Published 3 June 2024

Keywords:
ARDL
LSTM
crude palm oil
forecasting

DOI:
10.24191/jeeir.v12i2.2787

ABSTRACT

Understanding the volatile nature of palm oil prices is essential due to its profound economic and market implications. The complexity of forecasting palm oil prices stems from the interplay of various demand and supply forces, making it challenging for scholars to pinpoint the key determinants. This study addresses the intricate challenges in the palm oil industry, including price volatility, shifting consumer preferences, and environmental sustainability, by analysing the factors influencing Malaysian Crude Palm Oil (CPO) pricing dynamics. Utilising data from the Malaysian Palm Oil Board, covering the period from January 2004 to December 2021, we examined the impact of these variables on CPO prices. Methodologically, we employed Autoregressive Distributed Lag (ARDL) and Long Short-Term Memory (LSTM) models to evaluate and forecast CPO prices. Our findings indicate that the LSTM method outperformed the ARDL method in terms of forecasting accuracy. Specifically, the LSTM model showed superior performance when using a selection of ten independent variables identified through LASSO and SHAP estimation, compared to using eleven or four variables based on ARDL regression results. The analysis underscores the significant influence of weather conditions and macroeconomic factors, particularly tax rates, on CPO prices. These findings contribute to a deeper understanding of market dynamics and enhance the accuracy of CPO price forecasting.

1. Introduction

Understanding the pricing patterns of Crude Palm Oil (CPO) is essential due to their significant impact on numerous sectors. Analysing these price changes is crucial for stakeholders, including government bodies, farmers, investors, and palm oil production companies, as it aids in maintaining an equilibrium

[*] Corresponding author. *E-mail address*: imbar074@uitm.edu.my

between supply and demand in the broader market where palm oil is a key commodity (Oosterveer, 2015; Md Isa et al., 2016; Murphy et al., 2021).

Researchers have highlighted the value of understanding core economic principles to grasp price fluctuations. The Cobweb Theorem, described by Pashigian (2008), explains the cyclical nature of supply and demand in markets with production response delays. This theory helps in understanding market price fluctuations resulting from disparities in production and consumption. Such insights enable the development of effective strategies informed by historical data to modify supply and demand patterns. Additionally, price determination varies across different markets. In stock markets, prices are set through buyer and seller interactions, reflecting their perceptions and reactions to market conditions (Peterson, 2014; Ma et al., 2021). Conversely, commodity pricing, including CPO, involves diverse methods such as spot markets, futures markets, and direct transactions.

Comprehending CPO pricing is complex, requiring consideration of various factors like currency value fluctuations, trade restrictions, weather conditions, population growth, and the cost of other commodities (Chandrarin et al., 2022; Cespedes & Velasco, 2012; Enghiad, Headey & Fan, 2008; Wilson & Cacho, 2007 Zainalabidin & Rahim, 2012; Zaidi et al., 2021). The demand dynamics are also significantly influenced by the pricing of alternative vegetable oils and broader economic patterns. This complexity hinders accurate forecasting of future trends, affecting the reliability of predictions.

Given this context, accurate methods for predicting CPO prices must account for the simultaneous changes in these multifaceted determinants. Moreover, unforeseen events or changes in any single factor can lead to significant deviations from expected prices. This underscores the need for models that are comprehensive in incorporating a wide range of variables and adaptable to sudden shifts in market dynamics. The challenge lies in creating predictive models that can navigate the complexity of the factors at play and provide reliable forecasts in the face of uncertainty.

Recent discourse has increasingly focused on comparing traditional statistical methods like the ARDL approach, which explores complex correlations between various parameters and CPO prices, with modern AI techniques such as the LSTM model. Traditional approaches, such as ARDL, have been fundamental in analyzing long-term relationships between variables. However, with the advent of advanced computational methods, there has been a shift towards utilizing models like Artificial Neural Networks (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS). These models have shown superior performance in capturing non-linear patterns in data. For instance, a study by Karia et al. (2013) demonstrated that ANN models outperformed both ANFIS and ARFIMA models in predicting CPO prices (Karia et al., 2013).

However, despite their advantages, ANN and ANFIS models still face challenges such as overfitting and difficulty in understanding the rules generated by the models. These limitations have led to the development of more sophisticated methods like Long Short-Term Memory (LSTM) networks. LSTM models have been found to be particularly effective in time-series forecasting due to their ability to capture long-term dependencies in data. A study by Zhang and Hong (2022) highlighted the superior accuracy of LSTM models compared to ANN and ARIMA models in forecasting crude oil prices, suggesting that LSTM can also be effectively applied to CPO price forecasting (Zhang & Hong, 2022).

The evolution from traditional statistical methods to advanced AI models represents a significant shift in predictive modeling. While traditional models like ARDL provide a solid foundation, the integration of AI techniques, particularly LSTM, offers enhanced accuracy and the ability to manage complex, non-linear data patterns more effectively.

The ARDL approach is pivotal for identifying significant variables that influence CPO prices. The analysis involves investigating the short-term and long-term connections between CPO prices and a group of independent factors. The study used the LSTM method, renowned for its ability to manage complex temporal relationships in time-series data, to predict future CPO prices after identifying key factors. The LSTM method's capacity to learn from sequences of data and their underlying patterns makes it an exceptional tool for forecasting scenarios characterised by intricate interactions of various factors over time. By focusing on the most influential factors identified by the ARDL method, the LSTM method can potentially enhance the precision of forecasts. This synergistic approach between the ARDL and LSTM methods allows for a more refined analysis, where the ARDL method's strength in variable identification lays the groundwork for the LSTM method to leverage this information in making accurate future CPO price predictions (Hamid & Shabri, 2017; Ofuoku & Ngniatedema, 2022). Hence, this study's aim is to contrast the ARDL methodology with the LSTM method, an AI-based technique adept at detecting intricate data patterns over time. Our research focuses on selectively choosing factors that are crucial for improving the accuracy of forecasting models. The procedure is improved by using the Least Absolute Shrinkage and Selection Operator (LASSO) test in the ARDL framework to identify a subset of important variables. Such meticulous variable selection showcases the importance of precise variable identification in navigating the complex dynamics of CPO pricing, aiming to achieve the highest precision in prediction outcomes and offer strategic insights for decision-making in the CPO sector and related industries (Lu et al., 2021).

In summary, by synergising the ARDL and LASSO approach's variable selection process with the LSTM method's advanced pattern recognition capabilities, this research aims to forge a path to the most accurate CPO price predictions possible. Our research seeks to enhance understanding of the determinants of CPO, forecasting pricing trends, and guiding the industry towards a promising future. By utilising established economic concepts with advanced statistical and AI forecasting methods, this study aims to comprehensively understand the CPO market's intricacies, contributing to a dynamic and informed future in the sector.

2. Literature review

Palm oil holds immense importance as an agricultural product in Malaysia, substantially contributing to the country's economy. However, the palm oil industry faces challenges, including fluctuating pricing, changing consumer preferences, and environmental concerns. To address these issues, understanding the factors influencing short- and long-term palm oil pricing in Malaysia is crucial. This literature review aims to provide an overview of the available research on the determinants of Malaysian CPO prices.

The discovery of prices plays a fundamental role in the functioning of financial markets, where market players determine the pricing of assets such as stocks, bonds, and commodities based on supply and demand dynamics (Tomek & Kaiser, 2014; Netayarak, 2007). This process ensures that prices accurately reflect the underlying value of the traded items. The price discovery mechanism varies depending on the market and asset being traded. For example, in the case of stocks, price discovery occurs through transactions between buyers and sellers on the stock market, reflecting their valuations and prevailing market conditions. Similarly, price discovery occurs through spot markets, futures markets, and over-the-counter transactions for commodities like CPO. Buyers and sellers negotiate prices based on factors such as current supply and demand, production costs, and broader economic conditions.

At both the local and international levels, CPO prices are influenced by the interplay of supply and demand forces. Key factors that influence CPO prices include production and stock levels of CPO and its alternatives (e.g., soybean oil), exchange rates (as a measure of competitiveness and trade), and global economic uncertainties (Putri et al., 2019; Zaidi et al., 2021). Economic considerations significantly affect Malaysian CPO prices, as the demand for CPO is influenced by global economic indicators such as exchange rates, GDP, and the Consumer Price Index (CPI) (Putri et al., 2019; Zaidi et al., 2021).

CPO export taxation also plays a crucial role in price dynamics. Research has explored the effects of export taxes on CPO prices, with findings suggesting that such taxes may impact the competitiveness of CPOs on the global market, leading to price fluctuations (Hisham et al., 2019). Implementing an import tax on CPO in Indonesia, for instance, resulted in decreased demand and price declines, with significant implications for the CPO industry (Hisham et al., 2019). Conversely, reductions in CPO export taxes in Malaysia led to increased demand and price rises, fostering economic benefits (Amin et al., 2024).

https://doi.org/10.24191/jeeir.v12i2.2787

Moreover, CPO production levels, particularly in Indonesia, have influenced worldwide CPO price changes, reflecting the global impact of production trends. A rise in CPO prices may also adversely affect the price of soybean oil and export and production in Malaysia (Hassan & Balu, 2016).

In light of CPO's economic significance, accurate forecasting of CPO prices is essential in both the agricultural and financial sectors. Recent research has explored the ARDL methods and LSTM neural networks to forecast and interpret multivariate CPO price predictions. ARDL The Autoregressive Distributed Lag (ARDL) approach to forecasting has gained prominence for its versatility in econometric analysis. This method, adept at capturing both short-term dynamics and long-term equilibrium relationships, is particularly beneficial in financial and economic forecasting due to its flexibility with variables of different integration orders (Pesaran et al., 2001). A significant advantage of ARDL lies in its robustness in small sample sizes, making it a preferred choice in studies with limited data (Pesaran & Shin, 1995). However, challenges arise in method complexity and sensitivity to specification, necessitating careful selection of variables and lag lengths (Pesaran et al., 2001). Recent advancements integrate ARDL with machine learning techniques, enhancing forecasting accuracy and method robustness (Du et al., 2020).

Studies have highlighted the potential of LSTM methods in analysing complex relationships in multivariate time series data (Sagheer, A., & Kotb, M.; 2019), making it a suitable method for forecasting CPO prices influenced by various economic and environmental factors. Furthermore, multivariate LSTM methods have been utilised in financial time series prediction, addressing the challenges posed by global market dynamics and macroeconomic variables that impact CPO prices (Widiputra et al., 2021; Urolagin et al., 2021). The applicability of LSTM methods in long-term forecasting aligns well with the dynamic nature of CPO prices, where long-term trends and various influencing factors play significant roles in determining its price (Althelaya et al., 2018). The effectiveness of LSTM-FCNs in multivariate time series classification has also been emphasised, holding promise in classifying different CPO price patterns and identifying potential market trends (Karim et al., 2019). In conclusion, the ARDL and LSTM methods offer the potential for enhanced forecasting of CPO prices, enabling more accurate and interpretable predictions. Understanding the factors influencing CPO pricing is critical for navigating the complexities of the market. Further research to optimise the integration of ARDL and LSTM methods for CPO price forecasting is expected to yield even more accurate and insightful results, supporting better decision-making for investors and policymakers in the agricultural and financial sectors. The combination of economic factors and advanced forecasting techniques is key to unlocking a comprehensive understanding of CPO price movements, supporting the sustainable growth of the CPO industry in Malaysia and beyond.

3. Data and methodology

This study utilises monthly frequency data from January 2004 to December 2021, aiming to develop an Autoregressive Distributed Lag (ARDL) method incorporating eleven variables, with the Crude Palm Oil (CPO) price as the dependent variable. The independent variables in the model include the Export of CPO, CPO Production, CPO Export Tax, Stock of CPO, Rainfall (representing weather), Population, Economic Growth, Global Consumption, Price of Soybean, Price of Sunflower, Exchange Rate, and the Consumer Price Index.

However, this study acknowledges that the exclusion of certain variables may limit the comprehensiveness of the analysis. For instance, factors such as geopolitical events, which can significantly impact commodity prices, were not explicitly modeled. The choice of variables was also influenced by data availability and quality, leading to the exclusion of potentially relevant but less accessible data.

To ensure the reliability and accuracy of our findings, data was sourced from various reputable institutions, such as the Malaysian Palm Oil Board, Refinitiv, and Bloomberg. These sources were selected for their comprehensive and precise data on the relevant variables, enhancing our research conclusions' robustness and credibility.

https://doi.org/10.24191/jeeir.v12i2.2787

In addition to the traditional ARDL method, this research introduces the Long Short-Term Memory (LSTM) forecasting method. LSTM, a recurrent neural network (RNN), is particularly adept at handling time series data. Its ability to capture temporal dependencies and patterns in the dataset makes it a valuable tool for our analysis. By employing the ARDL method and LSTM, the study seeks to explore complex relationships within the data and potentially improve the accuracy of predictions regarding CPO prices. This dual-method approach allows for a more nuanced understanding of the factors influencing CPO pricing, contributing significantly to the field. ARDL forecasting uses Eviews to analyse and forecast the variables, and LSTM forecasting uses Python.

Table 1: Variables description

Notation	Variable	Description
CPO	CPO Price	Determined by supply-demand factors, production costs, market conditions, and trade dynamics.
POE	Palm Oil Export	The volume of international sales is driven by global demand, trade policies, and market competition.
POP	Palm Oil Production	Influenced by factors such as weather, technology, pests, and government policies.
POS	Palm Oil Stock	Based on production, trade volumes, market demand, and storage capacities.
TR	Tax Rate	Rate affecting palm oil's profitability, price, and market competition.
W	Weather	Climatic conditions influencing production, yield, and pest incidence.
P	Population	Determines demand based on size and consumption patterns.
SBP	Soybean Price	Affects palm oil pricing as a competing food and industrial ingredient.
SFP	Sunflower Price	Influences palm oil pricing as an alternative oilseed.
EG	Economic growth	Impacts price through income, spending, and trade dynamics.
ER	Exchange Rate	Affects global competitiveness and trade pricing.
CPI	Consumer Price Index	Indicator of inflation impacting production costs and purchasing power.

3.1 ARDL method

As mentioned by Pesaran et al. (2001), the advantage of ARLD is that the variables can be estimated with the combination of I (0) and I (1) series at the same time, with the single equations setup, that makes it simple to implement and interpret. This paper stabilises the variance of the series of all variables transformed into logarithmic form. Below is the ARDL method that used in this study:

$$\begin{aligned} CPO_t = \alpha &+ \sum_{i=1}^{p} \beta_i CPO_{t-i} + \sum_{j=0}^{q} \gamma_j POE_{t-j} + \sum_{k=0}^{r} \delta_k POP_{t-k} + \sum_{l=0}^{s} \theta_l POS_{t-l} \\ &+ \sum_{m=0}^{u} \lambda_m TR_{t-m} + \sum_{n=0}^{v} \mu_n W_{t-n} + \sum_{o=0}^{w} \nu_o P_{t-o} + \sum_{p=0}^{x} \xi_p SBP_{t-p} \\ &+ \sum_{q=0}^{y} \rho_q SFP_{t-q} + \sum_{r=0}^{z} \sigma_r EG_{t-r} + \sum_{s=0}^{a} \varphi_s ER_{t-s} + \sum_{t=0}^{b} \psi_t CPI_{t-t} \\ &+ \epsilon_t \end{aligned} \tag{1}$$

Where:
α = Intercept term
$\beta_i, \gamma_j, \delta_k, \theta_l, \lambda_m, \mu_n, \nu_o, \xi_p, \rho_q, \sigma_r, \varphi_s, \psi_t$ = Coefficients of the respective lagged terms

$p, q, r, s, u, v, w, x, y, z, a, b$ = Maximum lags for each variable
ϵ_t = Error term

ARDL methods are versatile statistical tools for analysing and forecasting complex interactions among multiple time series variables (Pesaran et al., 2001; Banerjee et al.,1993) concisely offer the following capabilities:

I. Capture Dynamics: They can simultaneously method short-term fluctuations and long-term relationships between variables.

II. Handle Non-stationarity: These methods are adept at dealing with variables that are not stationary, making them highly flexible for various types of data.

III. Cointegration Analysis: Multivariate ARDL methods can test and estimate long-term equilibrium relationships among variables.

IV. Estimate Impacts: They allow for estimating immediate and long-term effects of changes in one variable on others.

V. Complex Interactions: The methods can incorporate multiple variables to analyse complex interactions and feedback mechanisms.

VI. Forecasting: They are powerful in forecasting future values based on historical data and variable interactions.

VII. Robust and Flexible: These methods provide robust results across different specifications and offer flexibility in including lagged variables.

3.2 Diagnostic tests

The stability of the ARDL method was ensured through a comprehensive evaluation. For initial stability, three tests were applied: Augmented Dickey-Fuller (ADF) identified unit roots, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) assessed stationarity, and Zivot-Andrews (ZA) detected structural breaks. Diagnostic tests were conducted, including Breusch-Godfrey LM for serial correlation, Cameron & Trivedi’s IM-test for heteroskedasticity, and Skewness/Kurtosis tests for normality. Distribution was confirmed using techniques like normal probability plots. Meanwhile, structural stability was evaluated using recursive residuals’ cumulative sum (CUSUM).

3.3 Long Short-Term Memory (LSTM) method

The Long Short-Term Memory (LSTM) network is utilised in this study to forecast palm oil prices. LSTM is a type of recurrent neural network (RNN) that is particularly effective at learning from time series data due to its ability to maintain and update a cell state over time, which captures long-term dependencies. The LSTM forecasting method is structured as follows:

Input Sequence:

$$X_t = [POE_t, POP_t, POS_t, TR_t, W_t, P_t, SBP_t, SFP_t, EG_t, ER_t, CPI_t] \tag{2}$$

LSTM Processing:

$$h_t = LSTM(X_t) \tag{3}$$

Forecasted Output:

$$CPO_{t+1} = W_y * h_t + b_y \tag{4}$$

Where:
X_t is the input sequence consisting of the independent variables at time t:

- POE_t : Export of CPO
- POP_t : Production of CPO
- POS_t : CPO Export Tax
- TR_t : Stock of CPO
- W_t : Weather
- P_t : Population
- SBP_t : Price of Soybean
- SFP_t : Price of Sunflower
- EG_t : Economic Growth
- ER_t : Exchange Rate
- CPI_t : Consumer Price Index

h_t is the hidden state at time t after processing the input sequence with the LSTM network.

CPO_{t+1} is the forecasted palm oil price at time t+1.

W_y and b_y are the weight matrix and bias term of the output layer.

In practice, the LSTM method architecture can be designed with multiple LSTM cells, dropout layers, and possibly other recurrent or dense layers to capture the temporal patterns and dependencies in the multivariate time series data.

Multivariate Long Short-Term Memory (LSTM) methods are advanced neural networks designed for forecasting tasks involving multiple interacting time series (Hochreiter, S., & Schmidhuber, J., 1997). The key capabilities include follows:

I. Handling Sequential and Multivariate Data: They excel at processing time series data with multiple input variables, capturing the dynamic interactions among them.

II. Modelling Long-term Dependencies: LSTMs can remember and utilise long-term historical information, which is crucial for predicting future trends based on past data.

III. Learning Non-linear Relationships: These models are adept at identifying complex, non-linear patterns in data, surpassing traditional linear models in performance.

IV. Multi-step Forecasting: They can simultaneously predict several future time steps, which are practical for both short-term and long-term forecasting needs.

V. Automatic Feature Learning: LSTMs automatically learn relevant features from raw data, minimising the need for manual feature engineering.

VI. Robustness to Missing Data: Their recurrent structure makes them relatively resilient to gaps or missing values in time series data.

3.4 Features selection

In this study, not all factor data were significantly associated with CPO prices. To address this, the study initially utilized the LASSO method to remove irrelevant factors. This helped the researcher to obtain a more focused set of important variables for analysis. To further refine the selection, this study employed Random Forest to rank the feature importance, enabling the identification of the most influential factors. The core principle of LASSO is to compress the coefficients of irrelevant variables to zero in the regression problem (Tibshirani 1996). The regression problem in this study is expressed as follows:

$$y_{\mathrm{i}} = \omega^{\mathrm{T}} x_{\mathrm{i}} + b \tag{5}$$

$$J(w) = \frac{1}{m}\sum_{i=}^{m}(y_{\mathrm{i}} - \omega^{\mathrm{T}} x_{\mathrm{i}})^2 + \lambda \sum_{i=1}^{m}|w_{\mathrm{i}}|, \lambda > 0 \tag{6}$$

Where y_i, x_i and ω^T represent the monthly CPO prices, factors, and regression coefficients. The cost function J(w) is introduced to evaluate the accuracy of the regression method. It is necessary to find the λ ¬ that minimises the value of J(w).

3.5 Division of train set and test set

For all models, 216 months of data from January 2004 to December 2019 were utilised as the training set. The test set included data from January 2020 to December 2021. It was used to verify predictions made 6, 12, and 24 months ahead starting from January 2020.

3.6 Model accuracy test

The model selection involved using four error measures to evaluate the accuracy of the employed models. In particular, we adopted two scale-dependent metrics, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used with two scale-independent metrics: Mean Absolute Percentage Error (MAPE) and symmetric Mean Absolute Percentage Error (sMAPE).

4. Result

4.1 Variables selection

This research study explores the efficacy of Multivariate Autoregressive Distributed Lag (ARDL) and Long Short-Term Memory (LSTM) methods in forecasting Crude Palm Oil (CPO) prices, incorporating a set of 11 macroeconomic variables. Initial analyses involved conducting stationarity tests for each variable to determine their integrated order. Variables CPO, POP, and TR were found to be integrated of order 1 (I(1)), whereas variables POE, POS, W, P, SBP, SFP, EG, ER, and CPI exhibited stationarity at level (I(0)). This diversity in integration orders indicates the presence of both short-term dynamics and long-term relationships within the model.

Following the stationarity identification, we selected relevant factors based on the ARDL method, specifying the data using the Akaike Information Criterion (AIC) and considering a maximum of four lags for the dependent and independent variables. The lag structure for each variable was determined as ARDL (2, 4, 0, 2, 0, 0, 2, 0, 3, 0, 0, 0) for the 10-variable model and ARDL (2, 0, 2, 0, 3) for the 4-variable model.

https://doi.org/10.24191/jeeir.v12i2.2787

Further refinement involved employing the Least Absolute Shrinkage and Selection Operator (LASSO) test, aiding in identifying the most influential variables for accurate forecasting. The LASSO test excluded one variable from the initial set, resulting in a refined set of 10 critical factors. The coefficient compression and its impact on model performance are illustrated in Figure 1.

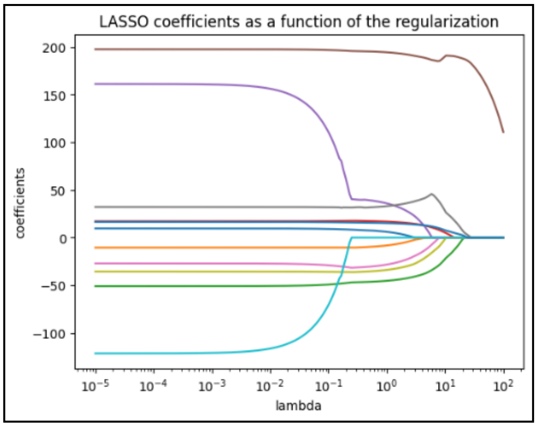

Figure 1. LASSO Coefficients

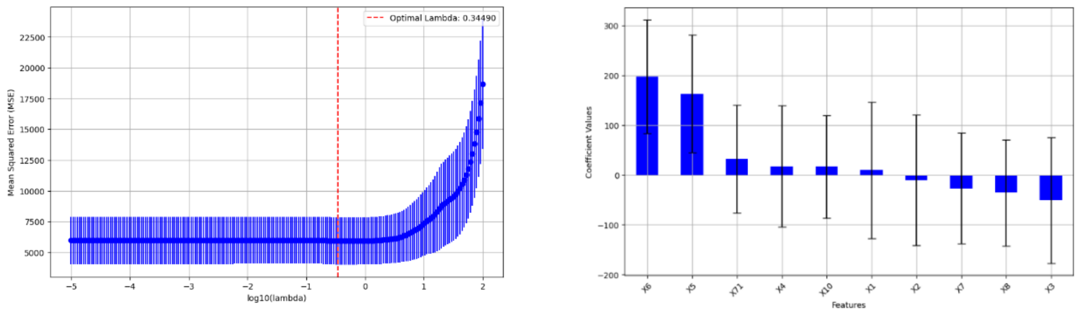

Figure 1. (left) MSE vs Log10(Lambda) with Standard Errors - Lasso; (right) Top 10 Lasso Selected Coefficients with Error Bars

Table 2: Variables selection

Number of IV	Method	Variables
11	All of the IV	POE, POP, POS, TR, W, P, SBP, SFP, EG, ER, CPI
10	LASSO	POE, POP, POS, TR, W, P, SBP, SFP, EG, CPI
4	ARDL	POP, POS, SBP, SFP

https://doi.org/10.24191/jeeir.v12i2.2787

4.3 ARDL multivariate forecasting

Table 2: Result of ARDL forecasting

No. of Independent Variables	RMSE	MAE	MAPE (%)	SMAPE(%)
11	0.1798	0.1434	218.3598	109.4485
10	0.1799	0.1439	217.6802	110.8776
4	0.1928	0.1512	214.9145	111.3982

The analysis of the ARDL method's forecasting performance reveals that the method with 11 independent variables delivers the most accurate predictions compared to configurations with 10 or 4 variables. This indicates the ARDL method's capability to effectively capture the underlying relationships within the data, leading to more precise forecasts.

A detailed examination of the performance metrics—namely Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE)—was conducted to evaluate the method's accuracy across different configurations. Lower values in these metrics reflect better performance, indicating smaller discrepancies between the predicted and observed values.

For the ARDL method with 11 independent variables, the results show an RMSE of 0.1798, an MAE of 0.1434, a MAPE of 218.3598%, and a SMAPE of 109.4485%. When the number of independent variables is reduced to 10, the method exhibits similar error metrics, with an RMSE of 0.1799, an MAE of 0.1439, a MAPE of 217.6802%, and a SMAPE of 110.8776%. The ARDL method with only 4 independent variables, however, produces higher error metrics, with an RMSE of 0.1928, an MAE of 0.1512, a MAPE of 214.9145%, and a SMAPE of 111.3982%.

These results suggest that while the ARDL method with 11 independent variables offers the best accuracy, the performance difference between the methods with 11 and 10 variables is marginal. However, a noticeable decline in accuracy occurs when the number of independent variables is reduced to 4, highlighting the importance of including a sufficient number of relevant variables for optimal forecasting.

In conclusion, the ARDL method with 11 independent variables is identified as the most effective configuration for forecasting in this context. The marginal differences between the 11-variable and 10-variable methods suggest that either could be suitable depending on specific forecasting requirements. The choice of method should consider the trade-off between method complexity and accuracy, with further refinement possible through additional statistical analyses or validation techniques to enhance the robustness of the forecasting outcomes.

4.4 LSTM multivariate forecasting result.

Table 3 Forecast performance of the multivariate LSTM method with different size factor sets

No. of Independent Variables	RMSE	MAE	MAPE (%)	SMAPE(%)
11	0.0508	0.0395	165.4761	71.1603
10	0.2488	0.2091	75.2272	55.9404
4	0.0516	0.0354	173.0183	79.6388

The analysis of the LSTM results indicates that a model utilising 11 independent variables demonstrates superior forecasting accuracy compared to configurations with 10 or 4 variables. This outcome underscores the LSTM model's capacity to effectively capture intricate patterns within the dataset, resulting in more precise predictions.

A detailed evaluation of performance metrics—including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE)—was conducted across various combinations of independent variables. These metrics are critical in assessing model performance, with lower values indicating reduced discrepancies between the model's forecasts and the actual observed values.

The results of the LSTM model incorporating 11 independent variables reveal an RMSE of 0.0508, an MAE of 0.0395, a MAPE of 165.4761%, and a SMAPE of 71.1603%. In comparison, the model with 10 independent variables exhibited slightly higher RMSE and MAE values of 0.0516 and 0.0354, respectively, alongside MAPE and SMAPE values of 173.0183% and 79.6388%. The model utilising 4 independent variables demonstrated the highest error metrics, with an RMSE of 0.0533, an MAE of 0.0407, a MAPE of 175.6117%, and a SMAPE of 93.2599%.

These findings suggest that the LSTM model with 11 independent variables offers the most accurate forecasts, as evidenced by its lower error metrics relative to the models with 10 or 4 variables. However, the differences in performance between the models with 11 and 10 variables are minimal, indicating that the model with 10 variables also performs at a competitive level.

In conclusion, based on a comprehensive assessment of the performance metrics, the LSTM model with 11 independent variables emerges as the optimal choice for forecasting in this context. The selection of this model should be guided by considerations of model complexity, accuracy, and interpretability. Further refinement of the model may be achieved through additional statistical tests or cross-validation, ensuring robustness and reliability in forecasting outcomes.

The SHAP approach is used with the multivariate single-step LSTM method as an example for model interpretation. A single sample from January 2020 was selected for the local explanation, and the results are shown in Figure 2.

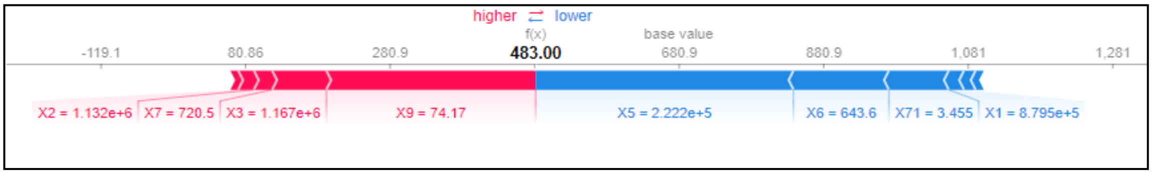

Figure 2. Impact of single sample characteristics (January 2020 forecast)

The 12 test set samples, from January 2020 to December 2021, were selected for the CPO price explanation, and the feature density scatter plot was drawn as shown in Figure 5. Each row corresponds to a feature, while the horizontal axis is the SHAP value. High and low feature values are indicated by red and blue, respectively.

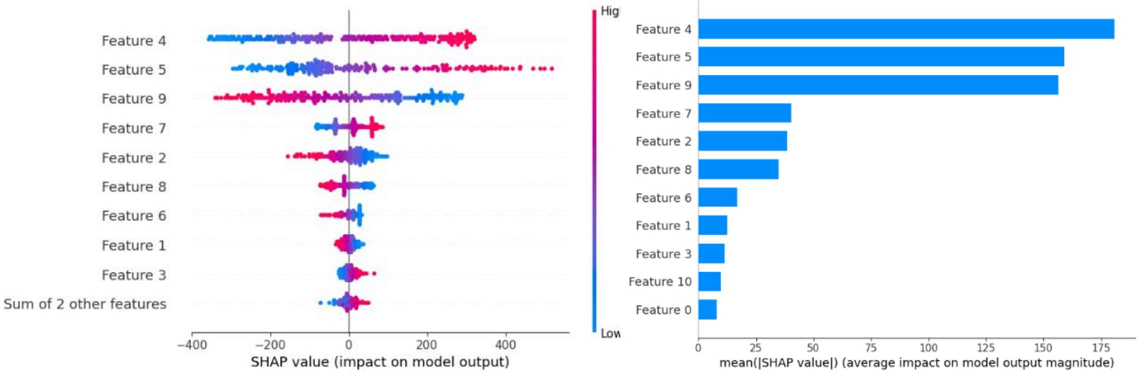

Figure 3. (left) Scatter plot of feature density; (right) Feature importance SHAP values

Specifically, futures 4, which is the tax rate, shows that its lower value will drive up the predicted value of CPO Prices. The feature 5, weather, indicates that its higher value pushes up the predicted value. The scatter of the remaining features oscillates about the SHAP value of 0. There is no spread to either side, which shows that these features have a lower correlation with the projected values. The absolute value of the SHAP value was initially calculated and then averaged to determine the feature significance, as shown in Figure 3.

4.5 Comparative analysis

The comparative analysis of the Multivariate ARDL and LSTM methods reveals that both approaches exhibit commendable forecasting accuracy. The LSTM methods' strengths lie in capturing complex temporal dependencies and non-linear patterns, while the Multivariate ARDL methods excel in assessing long-term relationships among the variables.

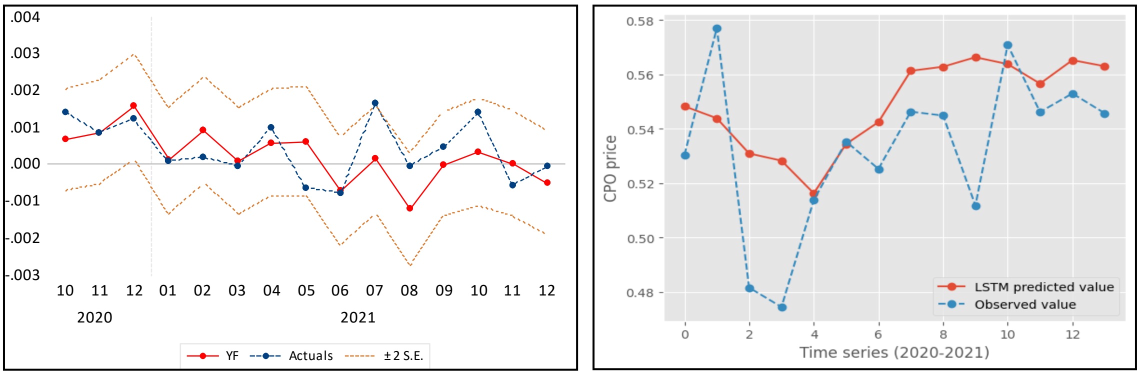

Figure 4. (left) Multivariate ARDL method; (right) Multivariate LSTM method

LSTM method significantly surpasses the ARDL method in forecasting accuracy, as evidenced by its lower RMSE, MAE, MAPE, and SMAPE values across various scenarios in this study. This superiority indicates that the LSTM method is more adept at providing precise forecasts for the specific context examined. In terms of consistency, the LSTM method demonstrates a more stable performance across different configurations of independent variables, unlike the ARDL method, which shows higher error metrics across the board. This contrast underscores the importance of selecting a forecasting method that aligns with the data's characteristics and the goals of the forecasting task.

The comparative analysis of the ARDL and LSTM models reveals notable differences in forecasting accuracy across various scenarios. The LSTM model consistently outperformed the ARDL model in terms of RMSE and MAE when utilizing a larger set of variables (11 and 10 variables). Specifically, the LSTM model's superior performance can be attributed to its ability to capture complex temporal dependencies and non-linear relationships inherent in time-series data. The LSTM architecture is designed to retain information over extended periods, making it particularly effective in scenarios where the data exhibits long-term patterns and correlations that traditional econometric models like ARDL may not fully capture.

However, the ARDL model demonstrated relatively strong performance when the number of variables was reduced to four. This could be explained by the ARDL model's strength in handling smaller datasets and its ability to model both short-term and long-term relationships between the dependent variable and a limited set of independent variables. The ARDL model is also less prone to overfitting in such cases, as it does not require extensive hyperparameter tuning like the LSTM model. This indicates that while the LSTM model is generally more powerful in handling complex datasets, the ARDL model may be more suitable for simpler scenarios where fewer variables are involved or when the dataset is relatively small.

4.6 Diagnostic test

Table 4. Forecast performance of the multivariate LSTM method with different size factor sets

Diagnostic test	**X^2 (P-value)**	**Result**
Breusch-Godfrey LM	0.95	No evidence of serial correlations
Breusch-Pagan-Godfrey	0.44	No evidence of heteroscedasticity
Ramsey RESET test	0.33	The model is specified correctly.

The diagnostic tests suggest that the model is stable and well-specified, with no signs of serial correlations or heteroscedasticity, as confirmed by the stability of the error correction coefficients shown in Figures 5.

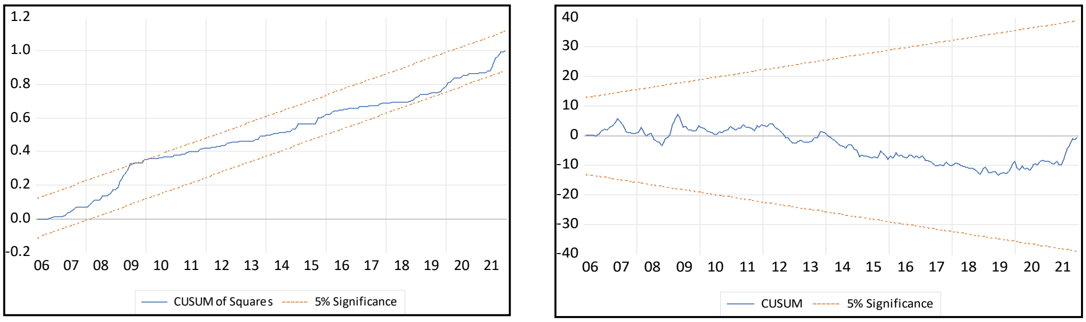

Figure 5. (left) CUSUM Test; (right) CUSUM of Squares test

https://doi.org/10.24191/jeeir.v12i2.2787

5. Discussion

The comprehensive analysis and comparison of the ARDL and LSTM methods for CPO price forecasting have provided valuable insights into their respective forecasting capabilities. This discussion section aims to delve deeper into the implications of the results obtained and their significance for decision-makers, researchers, and practitioners in the field of CPO price prediction.

One of the key takeaways from this study is the importance of variable selection in enhancing the accuracy of forecasting methods. The ARDL method's performance underscored the critical role of carefully selecting variables. In this study, the variables were selected based on the regression results, with 11 variables showing the best performance, followed closely by 10 variables. The results indicated that focusing on these relevant factors led to more accurate forecasts. This finding supports the notion that including too many or irrelevant variables can introduce noise and lead to overfitting, which diminishes the model's forecasting effectiveness (Castle, 2021). For decision-makers and researchers, this highlights the value of adopting robust variable selection techniques, such as the LASSO test, to improve the performance of multivariate forecasting models in CPO prices.

On the LSTM side, the analysis revealed that a model incorporating 11 independent variables consistently outperformed other configurations. This outcome underscores the significance of feature engineering and model parsimony. The LSTM method, known for its ability to capture complex temporal relationships and non-linear patterns, was most effective when a balanced number of variables were included. Models that strike a balance between incorporating relevant factors and avoiding unnecessary complexity tend to yield superior results.

The study also explored the impact of different variable configurations on forecasting accuracy, particularly in the context of the LSTM method. The analysis showed that, across all configurations, the LSTM model with 11 independent variables consistently outperformed other scenarios. This suggests that the selected variables, in conjunction with the LSTM architecture, are robust and adaptable to various data scenarios.

The ability to forecast accurately using different methods and variable sets is essential for decision-makers in the CPO market, where accurate short-term and long-term planning is crucial. Understanding the optimal model configuration and variable set for robust forecasting allows stakeholders to make more informed decisions regarding inventory management, production planning, and risk mitigation.

The trade-off between model complexity and performance is a critical consideration in model selection. While the LSTM method with 11 independent variables demonstrated superior accuracy, it is essential to recognize that increasing model complexity can raise interpretability challenges and computational demands. Therefore, selecting the optimal model should also consider contextual requirements, available computational resources, and the need for transparency in decision-making processes.

This study's use of performance metrics to evaluate the ARDL and LSTM methods provides a valuable framework for understanding the strengths and limitations of each approach. For decision-makers and scholars, these insights can inform the selection of forecasting methods best suited to their specific needs, ensuring that the chosen model delivers both accuracy and interpretability.

In conclusion, the results of this study highlight the importance of carefully selecting the forecasting method and variable configuration that best aligns with the data characteristics and forecasting goals. Whether using the ARDL or LSTM method, decision-makers should consider the balance between model accuracy, complexity, and interpretability to make informed decisions in the CPO market.

https://doi.org/10.24191/jeeir.v12i2.2787

6. Practical implications for forecasting

The results of this study have significant practical implications for real-world forecasting scenarios, particularly in the context of the palm oil industry. The superior performance of the LSTM model in most cases suggests that it is better suited for applications where the goal is to achieve the highest possible accuracy in forecasting CPO prices, especially when dealing with large and complex datasets. This makes the LSTM model a valuable tool for stakeholders who require precise forecasts for decision-making in areas such as inventory management, production planning, and pricing strategies.

On the other hand, the ARDL model's relatively strong performance with a smaller variable set highlights its utility in scenarios where data availability is limited, or when a simpler, more interpretable model is preferred. For instance, policymakers and industry analysts who require a straightforward and easily interpretable model for scenario analysis or policy evaluation may find the ARDL approach more suitable.

Furthermore, the differences in model performance underscore the importance of selecting the appropriate forecasting model based on the specific characteristics of the data and the forecasting objectives. In practice, this means that decision-makers should carefully consider the trade-offs between model complexity, interpretability, and accuracy. While LSTM may offer superior accuracy, it comes with increased computational demands and complexity, which might not be feasible in all situations.

Ultimately, this study suggests that a hybrid approach where both ARDL and LSTM models are employed depending on the context could offer a balanced solution, leveraging the strengths of each model to achieve more reliable and robust forecasts in the palm oil market.

7. Conclusion

This study focuses on developing reliable and accurate models for forecasting multivariate Crude Palm Oil (CPO) prices, highlighting the critical importance of selecting appropriate variables for prediction. The findings demonstrate the effectiveness of both the ARDL and LSTM methods, with the ARDL method consistently delivering superior accuracy and reliability across various scenarios. The research emphasizes the need to choose a forecasting method that is not only accurate but also straightforward and interpretable, ensuring it aligns well with the data's characteristics and the study's objectives. By exploring different variable configurations, the study underscores the necessity of balancing model complexity with performance. Additionally, the use of interpretability methods, such as SHAP, enhances our understanding of the key factors influencing CPO prices, with the ARDL-based variable selection proving essential in identifying these influential variables. This research provides valuable insights for decision-makers and researchers in the CPO industry, equipping them with the tools to make informed decisions in a dynamic market. However, the study also acknowledges certain limitations and emphasizes the need for ongoing research and data refinement to further enhance the accuracy of CPO price forecasting.

Acknowledgements

The authors would like to extend their heartfelt gratitude to Universiti Teknologi Mara (UiTM), Campus Shah Alam, and the Faculty of Business Management. Special thanks are due to the Faculty of Business Management at UiTM for their unwavering support, insightful critiques, and valuable suggestions, all of which have greatly influenced the direction and outcome of this study. The authors are also deeply appreciative of the patience, encouragement, and understanding shown by their families and friends throughout the research process. Their moral support has been an enduring source of strength and motivation.

https://doi.org/10.24191/jeeir.v12i2.2787

Conflict of interest statement

The authors hereby affirm that the conduct of this study has been executed without any commercial or financial relationships that could be construed as a potential conflict of interest. The integrity of the research process and results remain uncompromised, and there has been no influence from the funding sources that could be viewed as a conflict of interest.

References

Althaea, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018). Stock market forecast using multivariate analysis with bidirectional and stacked (LSTM, GRU). In 2018 *21st Saudi Computer Society National Computer Conference (NCC)* (pp. 1-7). IEEE. https://doi.org/10.1109/NCG.2018.8593076

Amin, S. I. M., Shukor, M. S., & Abdul-Rahman, A. (2024). Determinants of Malaysian palm oil export efficiency: A gravity stochastic frontier model. *Geografia-Malaysian Journal of Society and Space, 20*(2), 145-155.

Banerjee, A., Dolado, J., Galbraith, J. W., & Hendry, D. (1993). *Cointegration, error correction, and the econometric analysis of non-stationary data*. Oxford University Press. https://books.google.com.my/books?hl=en&lr=&id=LfVQEAAAQBAJ&oi=fnd&pg=PR7&dq=Banerjee,+A.,+Dolado,+J.,+Galbraith,+J.+W.,+%26+Hendry,+D.+(1993).+Cointegration,+error+correction,+and+the+econometric+analysis+of+non-stationary+data.+Oxford+University+Press&ots=-8SvWMGiUY&sig=twi7s6vceseLKcUE1oByDQ7ipPE

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2021). Selecting a model for forecasting. *Econometrics, 9*(3), 26. https://doi.org/10.3390/econometrics9030026

Cespedes, L. F., & Velasco, A. (2012). Assessing the impact of food trade policies on poverty alleviation and welfare. *Journal of Policy Modelling, 34*(6), 879-894. https://doi.org/10.1057/imfer.2012.22

Chandrarin, G., Sohag, K., Cahyaningsih, D. S., Yuniawan, D., & Herdhayinta, H. (2022). The response of exchange rate to coal price, palm oil price, and inflation in Indonesia: Tail dependence analysis. *Resources Policy, 77*, 102750. https://doi.org/10.1016/j.resourpol.2022.102750

Enghiad, A., Ufer, D., Countryman, A., & Thilmany, D. (2017). An Overview of Global Wheat Market Fundamentals in an Era of Climate Concerns. *International Journal of Agronomy, 2017*(1), 3931897.. https://doi.org/10.1155/2017/3931897.

Hamid, M. F. A., & Shabri, A. (2017). Palm oil price forecasting model: An autoregressive distributed lag (ARDL) approach. *AIP Conference Proceedings*, 1842(1). https://doi.org/10.1063/1.4982864.

Hassan, A., & Balu, N. (2016). Examining the long-term relationships between the prices of palm oil and soybean oil, palm oil production and export: Cointegration and causality. *Oil Palm Industry Economic Journal, 16*(1), 31-37. https://www.researchgate.net/publication/324983472

Headey, D., & Fan, S. (2008). Anatomy of a crisis: The causes and consequences of surging food prices. *Agricultural Economics, 39*(1), 375-391. https://doi.org/10.1111/J.1574-0862.2008.00345.X.

Hisham, A. A. M., Karim, Z. A., & Khalid, N. (2019). Determinants of capital expenditure spending in Malaysian palm oil industries: A dynamic panel data analysis. *Economic Journal of Emerging Markets*, 223-233. https://doi.org/10.20885/ejem.vol11.iss2.art9.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735-1780. https://doi:10.1162/neco.1997.9.8.1735

Karia, A., Bujang, I., & Ahmad, I. (2013). Forecasting on Crude Palm Oil Prices Using Artificial Intelligence Approaches. *American Journal of Operations Research, 3*(2), 259-267. https://doi.org/10.4236/AJOR.2013.32023.

Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2019). Multivariate LSTM-FCNs for time series classification. *Neural networks, 116*, 237-245. https://arxiv.org/pdf/1801.04503

Lu, Q., Sun, S., Duan, H., & Wang, S. (2021). Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model. *Energy Informatics, 4*(1), 1-20. https://doi.org/10.1186/s42162-021-00166-4

Ma, Q., Huang, J., Başar, T., Liu, J., & Chen, X. (2021). Reputation and Pricing Dynamics in Online Markets. *IEEE/ACM Transactions on Networking, 29*(4), 1745-1759. https://doi.org/10.1109/TNET.2021.3071506.

Md Isa, M. A., Samsudin, S., & Noh, M. K. A. (2016). Co-integration and causality between the spot and futures markets: pre and post implementation of NKEAs. Advances in Business Research International Journal, 2(1), 1-13.

Murphy, D.J., Goggin, K. & Paterson, R.R.M. (2021) Oil palm in the 2020s and beyond: challenges and solutions. CABI agriculture and bioscience, 2, 1-22. https://doi.org/10.1186/s43170-021-00058-3

Netayarak, P. (2007). *Agricultural Economics.* Bangkok: Thammasat University.

Ofuoku, M., & Ngniatedema, T. (2022). Predicting the Price of Crude Palm Oil: A Deep Learning Approach. *International Journal of Strategic Decision Sciences (IJSDS), 13*(1), 1-15. https://www.igi-global.com/article/predicting-the-price-of-crude-palm-oil/305830

Oosterveer, P. (2015). Promoting sustainable palm oil: viewed from a global networks and flows perspective. *Journal of Cleaner Production, 107,* 146-153. https://doi.org/10.1016/J.JCLEPRO.2014.01.019.

Pashigian, B. P. (2008). Cobweb Theorem: The New Palgrave Dictionary of Economics. *Springer*. doi, 10, 978-1. https://books.google.com/books?hl=en&lr=&id=EO40DAAAQBAJ&oi=fnd&pg=PR2&dq=+%22Cobweb+Theorem%22.+The+New+Palgrave+Dictionary+of+Economics+(2nd+ed.)&ots=8m56JYBDlX&sig=7-xm8vvetDXl5pqFeoolI2_XzbI

Pesaran, M. H., & Shin, Y. (1995). *An autoregressive distributed lag modelling approach to cointegration analysis* (Vol. 9514, pp. 371-413). Cambridge, UK: Department of Applied Economics, University of Cambridge.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics, 16*(3), 289-326. https://doi.org/10.1002/jae.616

Peterson, P. (2014). Fixing Prices and Fixing Markets. *farmdoc daily, 4*. https://ageconsearch.umn.edu/record/283094/files/fdd250614.pdf

Putri, P. Y., Achsani, N. A., & Pranowo, K. (2019). The Effects of Macroeconomic Variables and Corporate Financial Performance on Stock Prices of Palm Oil Companies in Indonesia. *Jurnal Manajemen & Agribisnis, 16(*1), 12-12. /DOI: http://dx.doi.org/10.17358/jma.16.1.12

Ricci, E., Peri, M., & Baldi, L. (2019). The Effects of Agricultural Price Instability on Vertical Price Transmission: A Study of the Wheat Chain in Italy. *Agriculture, 9*(2), 36. https://doi.org/10.3390/AGRICULTURE9020036.

Sagheer, A., & Kotb, M. (2019). Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems. *Scientific reports, 9*(1), 19038. https://www.nature.com/articles/s41598-019-55320-6

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tomek, W.G. & Robinson, K.L. 1981. Agricultural Product Prices. Ithaca, New York.: Cornell University Press. https://www.cabidigitallibrary.org/doi/full/10.5555/19811875109

Tomek, W. G., & Kaiser, H. M. (2014). *Agricultural product prices*. Cornell University Press.

Urolagin, S., Sharma, N., & Datta, T. K. (2021). A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting. *Energy, 231*, 120963. https://www.sciencedirect.com/science/article/pii/S0360544221012111

Widiputra, H., Mailangkay, A., & Gautama, E. (2021). Multivariate cnn-lstm model for multiple parallel financial time-series prediction. *Complexity, 2021* (1), 9903518. https://www.hindawi.com/journals/complexity/2021/9903518/

Wilson, N., & Cacho, J. (2007). Linkage between foreign direct investment, trade and trade policy: an economic analysis with application to the food sector in OECD countries and case studies in Ghana, Mozambique, Tunisia and Uganda. https://www.oecd-ilibrary.org/content/paper/152275474424

Zaidi, M. A. S., Karim, Z. A., & Zaidon, N. A. (2021). External and internal shocks and the movement of palm oil price: SVAR evidence from Malaysia. *Economies, 10*(1), 7. https://doi.org/10.3390/economies10010007

Zainalabidin, S. M., & Rahim, K. A. (2012). Impact of climate change on palm oil production. In *Proceedings of USM-AUT International Conference 2012 Sustainable Economic Development: Policies and Strategies* (Vol. 167, p. 109).

Zhang, K., Hong, M., Zhang, K., & Hong, M. (2022). Forecasting crude oil price using LSTM neural networks. *Data Sci. Financ. Econ, 2*, 163-180. https://doi.org/10.3934/dsfe.2022008.

About the authors

Mohd Shahrin, MSc is a PhD Student at Universiti Teknologi MARA (UiTM) and is an officer in the Department of Monitoring State Agencies and Investment at Ministry of Finance Sabah (ORCID: http://orcid.org/0009-0003-2290-9733). His main research interest is in Business Economics and Econometrics.

Imbarine Bujang, PhD, serves as the Deputy Rector (Academic) and is a Professor at the Faculty of Business and Management, Universiti Teknologi MARA (UiTM) Sabah, Malaysia (ORCID: http://orcid.org/0000-0002-1427-0963). He holds a PhD in Financial Econometrics from Edith Cowan University, Western Australia. His research primarily revolves around Financial Economics, Econometrics, Behavioural Finance, and Financial Management. He has been recognised for his significant contributions to these fields and has an extensive publication record. He is also an Independent Member and Chairman of the Audit and Compliance Committee at Saham Sabah Berhad since October 2021. His contact email is imbar074@uitm.edu.my.

Abdul Aziz Karia, PhD, is a Senior Lecturer in Business Economics at Universiti Teknologi Mara (UiTM) Sabah and an Associate Fellow at the Accounting Research Institute, HICoE, at the same university

(ORCID: http://orcid.org/0000-0002-4888-682X). His PhD in Economic Forecast was obtained from Universiti Teknologi MARA. Dr. Karia's primary research interests include economic forecasting and policymaking, with a focus on regional development. He has contributed to various academic publications and participates actively in state-level decision-making and policy papers. His contact email is Abdulaziz@uitm.edu.my.

Nurzahidah Baharudin, PhD, is a Senior Lecturer at the Faculty of Business and Management, Universiti Teknologi MARA (UiTM), Malaysia. She obtained her PhD from UiTM and specializes in the field of Finance. Her current academic responsibilities primarily involve teaching and learning. Her contact email is zahidah1002@uitm.edu.my.

Authors' contributions

Mohd Shahrin Bahar carried out the research, wrote the methodology, formal analysis, and wrote the original draft. Prof. Dr Bujang and Dr Aziz Karia conceptualised the central research idea and provided the theoretical framework. Prof. Dr Imbarine Bujang supervised the research progress. Dr Aziz Karia does supervision on methodology. Dr Nur Zahidah Baharudin anchored the review, revisions, and approved the article submission.