第 49 卷 第 3 期
2022 年 3 月

湖南大学学报（自然科学版）
Journal of Hunan University (Natural Sciences)

Vol. 49 No. 3
March 2022

Open Access Article https://doi.org/10.55463/issn.1674-2974.49.3.25

Modeling Vector Autoregressive and Autoregressive Distributed Lag of the Beef and Chicken Meat Prices during the Covid-19 Pandemic in Indonesia

Samingun Handoyo[1,2,3], Ying-Ping Chen[4,5], Tiara Mawidha Shelvi[1], Heni Kusdarwati[1]

[1] Department of Statistics, Brawijaya University, Malang 65145, Indonesia

[2] Department of Electrical Engineering and Computer Science-IGP, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

[3] Department of Electrical Engineering and Computer Science-IGP, National Chiao Tung University, Hsinchu 30010, Taiwan

[4] Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

[5] Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan

Abstract: The impact of the COVID-19 pandemic has spread to all aspects of life. Modeling the price of beef and chicken meat is very important for the government to avoid extreme fluctuations of both commodities in the prices so that society's purchasing power can be maintained. This study has several objectives, namely building VAR and ARDL models from multiple time series data (beef and chicken meat prices), conducting variable selection with forwarding subset selection on input lag in the ARDL model, and measuring the performance of the VAR and ARDL models on the both of beef and chicken meat prices based on the value of RMSE, MAE, and R_square both in the training and testing set. The novelty in this study is to propose an identification method for the lag inputs of the ARDL model based on the criteria of both the Alkaide Information criteria (AIC) value and the adjusted R square value by visualizing both criteria for all possible amounts of lag inputs. The results of the identification of the VAR model structure using the conventional method in time series modeling are yielded the different lag inputs that are compared to the ARDL model structure with lag inputs identified by using the proposed method. The ARDL model of the beef and chicken meat prices has better performance than the VAR model both on training and testing sets. In addition, the resulting VAR model also clearly shows the occurrence of overfitting problems.

Keywords: ARDL modeling, feature selection, multiple time series, VAR modeling.

印度尼西亚新冠肺炎大流行期间牛肉和鸡肉价格的向量自回归和自回归分布滞后建模

摘要：新冠肺炎大流行的影响已蔓延到生活的方方面面。模拟牛肉和鸡肉的价格对于政府避免两种商品价格的极端波动以维持社会的购买力非常重要。本研究有几个目标，即从多个时间序列数据（牛肉和鸡肉价格）构建向量自回归和自回归分布式滞后模型，在自回归分布式滞后模型中通过前向子集选择对输入滞后进行变量选择，并测量基于训练和测试集中均方根误差、平均绝对误差和R平方值的牛肉和鸡肉价格的向量自回归和自回归分布滞后模型。本研究的新颖之处在于，通过可视化所有可能的滞后输入量的两个标准，提出了一种基于碱信息标准值和调整后的R平方值标准的自回归分布式滞后模型的滞后输入的识别方法。在时间序列建模中使用传统方法识别向量自回归模型结构的结果产生了不同的滞后输入，并与

Received: 10 December, 2021 / Revised: 21 January, 2022 / Accepted: 10 February, 2022 / Published: 28 March, 2022

About the authors: Samingun Handoyo, Department of Statistics, Brawijaya University, Malang, Indonesia; Department of Electrical Engineering and Computer Science-IGP, National Yang Ming Chiao Tung University, Hsinchu, Taiwan; Department of Electrical Engineering and Computer Science-IGP, National Chiao Tung University, Hsinchu, Taiwan; Ying-Ping Chen, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan; Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan; Tiara Mawidha Shelvi, Heni Kusdarwati, Department of Statistics, Brawijaya University, Malang, Indonesia

使用该方法识别滞后输入的自回归分布式滞后模型结构进行了比较。牛肉和鸡肉价格的自回归分布滞后模型在训练和测试集上都比向量自回归模型具有更好的性能。此外，得到的向量自回归模型也清楚地显示了过拟合问题的发生。
关键词：自回归分布式滞后建模、特征选择、多时间序列、向量自回归建模。

1. Introduction

The machine learning modeling approach is output-oriented. There are two major approaches called descriptive and predictive modeling. Predictive modeling is divided into 2 types based on the response variable measuring scale. A regression model is built when a response variable has a numerical scale (interval or ratio), while a classification model is built when the response variable has a categorical scale (nominal or ordinal). A classification model has a primary goal to classify the unknown label of instances [1-3]. In statistics, regression modeling is more emphasized to explore the causality relationship between the response and predictors variable [4], but in machine learning, the regression modeling is oriented to capture all existing patterns in the data set to obtain a model that can predict accurately to the unknown value of response variable based on the known values of predictor variables [5-6].

Time series data are available abundantly in various fields. Each event has a certain pattern such as a trend, seasonal, or cycle. Past data can be used to identify patterns and then a time series model can be built to predict the unknown future value. Time series data involving only one observed variable is called univariate time series, while time-series data involving more than two variables is called multivariate (multiple) time series. Model development in univariate time series has been very sophisticated which produces hybrid models having very satisfactory performance, including hybrid models between wavelets and neural networks [7] and also modeling that hybridizes between SVM and LSTM [8]. However, multiple time series modeling faces many limitations including many constraints in modeling that cannot be met [9-10] and also unsatisfactory performance because they generally suffer from the overfitting problem [11].

In reality, few time series are interrelated to form multiple time series. Vector Autoregressive (VAR) is an autoregressive (AR) model developed in multiple time series. The implementation of the VAR model including had conducted by Marica and Horobet [12] which they selected a VAR model using a genetic algorithm. The stock price and exchange rate modeling by using the sign-restricted VAR model were carried out by Chen and Liu [13]. The identification of the VAR model structure is complex and the VAR model consists of multiple models having the same model structure has led to the development of a more flexible model called Autoregressive distributed lag (ARDL). Chandio et al [14] developed an ARDL model to predict agricultural production in Pakistan. Sohail et al [15] conducted a nonlinear ARDL analysis of air pollution and transport management in Pakistan. Although the ARDL model is a single model, identifying the ARDL model structure is also not an easy task. Feature selection in machine learning modelings such as those in Ircio et al [16] and Zhu et al [17] will be used to identify the structure of the ARDL model in this study.

The need for food is a primary need for the community for both plant and animal food needs. One of the most popular animal food commodities is beef and chicken. The demand for meat continues to increase every year also due to the increasing number of business actors in the culinary field. Zhang et al [18] have explored the factors influencing household meat purchases in China, while several researchers including Lusk et al [19] discussed changes in prices and marketing margins. Beef price predictions with time series models were carried out by Aguirre and Aguirre [20], and also by Zeng et al [21]. Meanwhile Alderiny et al [22] built a time series model to predict Saudi Arabia’s production and imports of broiler meat chickens.

In this study, the VAR and ARDL models were built from the prices of beef and chicken meat during the Covid19 pandemic in East Java, Indonesia. The forward subset feature selection method is used to identify the ARDL model structure of the two commodities and it is hoped that the right predictor features can be obtained so that the resulting ARDL model has a satisfactory performance. The model performance was evaluated based on the value of RMSE, MAE, and R square on both training and testing parts.

2. The Literature Review

Vector Autoregressive (VAR) can be considered as multi-response modeling with the predictor variable being autoregressive lag with order p from each of its constituent time series data which can be expressed as VAR(p). Meanwhile, Autoregressive Distributed lags (ARDL) is a multiple-time series modeling group with a single response variable and predictor variables in the form of a combination between autoregressive lags of order p and distributed lags of order q which can be expressed as ARDL(p,q). The distributed lag in this case is the lags formed from the explanation variable, namely the variable that causes or explains the variation in the response variable.

2.1. VAR Modeling

Identification of the VAR model structure is a very important step which includes stationarity test, determination of the optimum lag number, and Granger causality test. A stationary time series means that the time series has a zero mean and a constant variance. The selection of the optimum lag number and the Granger causality test aim to obtain the lag order number (p) from the time series data involved in the model structure [23]. The VAR model was obtained using ordinary least squares for estimating the parameters of the model structure [24].

Consider the process $\{y_t; t \in \mathbb{Z}\}$ follows a vector autoregressive model of order p, denoted $VAR(p)$ presented in equation (1) as the following [25]:

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,\ \ t \in \mathbb{Z} \quad (1)$$

where p is lag number (positive integer), A_i are fixed $(K \times K)$ coefficient matrices, $v = (v_1, \ldots, v_K)'$ is a fixed $(K \times 1)$ vector of intercept terms, and $u_t = (u_{1t}, \ldots, u_{Kt})'$ is a K-dimensional white noise with covariance matrix Σ_u that is assumed to be nonsingular. To make it easier to explain the conceptual framework of VAR modeling, in this discussion the VAR (1) model structure is used as a case study presented in equation (2) following

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

is equivalent to

$$\begin{aligned} y_{1t} &= v_1 + a_{11} y_{1t-1} + a_{12} y_{2t-1} + u_{1t} \\ y_{2t} &= v_2 + a_{21} y_{1t-1} + a_{22} y_{2t-1} + u_{2t} \end{aligned} \quad (2)$$

where $\operatorname{cov}(u_{1t}, u_{2s}) = \sigma_{12}$.

The model structures in equation (2) correspond to 2 regressions with different dependent (response) variables and identical explanatory variables.

The parameters of the VAR (1) model can be obtained by using the ordinary least squares (OLS) computed separately from each formula in equation (2.2). It is assumed that a time series $y_1 = [y_{11}, y_{21}]', \ldots, y_T = [y_{1T}, y_{2T}]'$ of the y variables is available. In addition, a pre-sample value $y_0 = [y_{10}, y_{20}]'$ is assumed to be available. Consider the first part of equation (2) as the following:

$$y_{1t} = v_1 + a_{11} y_{10} + a_{12} y_{2t-1} + u_{1t}; for$$
$$t = 1, \ldots, T$$
$$y_{11} = v_1 + a_{11} y_{10} + a_{12} y_{20} + u_{11}$$
$$y_{12} = v_1 + a_{11} y_{11} + a_{12} y_{21} + u_{12}$$
$$\vdots$$
$$y_{1T} = v_1 + a_{11} y_{1T-1} + a_{12} y_{2T-1} + u_{1T}$$

By defining of some terms as below

$$y_1 = [y_{11}, \ldots, y_{1T}]', X_1 = \begin{bmatrix} 1 & y_{1,0} & y_{2,0} \\ 1 & y_{1,1} & y_{2,1} \\ \vdots & \vdots & \vdots \\ 1 & y_{1,T-1} & y_{2,T-1} \end{bmatrix},$$

$\pi_1 = [v_{11}, a_{11}, a_{12}]'$, and $u_1 = [u_{11}, \ldots, u_{1T}]'$. Then the first part of equation (2.2) can be written as the following:

$y_1 = X\pi_1 + u$, and the OLS estimator of π_1 is given by

$$\hat{\pi}_1 = (X'X)^{-1} X' y_1 \quad (3)$$

2.2. Modeling ARDL with Stepwise Forward Feature Selection

An ARDL model refers to a model with lags of both the dependent and explanatory variables. An ARDL(p,q) model would have 1 lag on both variables which is expressed in equation (4) following [26]:

$$y_1 = \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 x_{t-2} + \cdots +$$
$$\alpha_{p+1} x_{t-p} +$$
$$\alpha_{p+2} y_{t-1} + \alpha_{p+3} y_{t-2} + \cdots + \alpha_{p+q+1} y_{t-q} + u_t \quad (4)$$

The order autoregressive q and order distributed p is obtained by using the forward stepwise feature selection which has a Procedure based on the idea that no predictor variables are in the model originally, but they are added one by one at a time. The selection procedure is [27-28]:

a. The first predictor selected to be entered into the model is the one with the highest correlation with the response. If the F statistic corresponding to the model containing this variable is significant (larger than some predetermined value), then the predictor is left in the model.

b. The second predictor examined is the one having the largest partial correlation with the response. If the F-statistic corresponding to the addition of this variable is significant, the predictor is retained.

c. The process is continued until all predictor variables are examined.

After all the predictor variables are entered into the model (as many as p models), then the model with the highest interpretability is selected. The criteria used as a measure in choosing the number of predictor variables that produce the best model are the Akaike's Information Criterion (AIC) and R square adjusted values. The best model has the lowest AIC value and the highest R square adjusted. The formula to calculate the AIC and R square adjusted values is

$$\text{AIC} = \frac{1}{n\hat{\sigma}^2}(\text{RSS} + 2d\hat{\sigma}^2) \quad (5)$$

$$\text{Adjusted R}^2 = 1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)} \quad (6)$$

where RSS is residual sum square, TSS is total sum square, *n* is the number of instances, *d* is the number of parameters [29].

2.3. The Measure Model Performance

The performance of the regression model is evaluated based on the accuracy of the model's predictions against the actual value. The measure of the accuracy of the regression model is generally based on the gap between the actual value and the predicted value. The correlation between the actual value and the predicted value squared is called R square or the coefficient of determination which in this case measures the level of accuracy of the predicted value generated by the model against the actual value. The value of R square (R^2) has a range of [0,1] where R^2 = 1 means that the regression model can predict the actual value with 100% accuracy or when presented in a plot of the actual versus the predicted values, the two coincide perfectly. Meanwhile, other measures of model performance include RMSE and MAE which provide a numerical value for the average model error in predicting the actual value [30]. RMSE gives a large weight to the predicted value with a large bias, but not so with MAE. Here is the formula to calculate MAE, RMSE, and R^2:

$$AE = \sum_{i=1}^{n} \frac{|\hat{Y}-Y|}{n}$$

$$RMSE = MSE^{1/2}, \quad \text{where} \quad MSE = \sum_{i=1}^{n} \frac{(\hat{Y}-Y)^2}{n}$$

$$R\ square = r^2,$$

where $r = \frac{n\sum \hat{Y}Y - \sum \hat{Y} \sum Y}{\sqrt{(n\sum \hat{Y}^2 - (\sum \hat{Y})^2)(n\sum Y^2 - (\sum Y)^2)}}$

where $\hat{Y}$ is the predicted value, and Y is the corresponding actual value.

3. The Data and Research Steps

The multiple time series studied in this study are the average price of beef and the average price of chicken meat at the consumer level in the province of East Java, Indonesia. The average daily prices of the two commodities are obtained based on the average daily price of the two commodities in 37 main markets in the East Java region. The data are provided by the East Java provincial industry office, which records basic commodity prices (kg) in 37 district/city wholesale markets. These two commodities are the main source of animal protein which is very important for the community and the government must be able to control price fluctuations so that the community is not harmed.

The development of the VAR and ARDL models is expected to be a tool to predict the prices of the two commodities so that the government can take policies to control prices in the event of very sharp (abnormal) price fluctuations. The stages of developing the VAR and ARDL models with a machine learning approach are as follows:

a. The data description is based on time series plot and summary statistics. It aims to get a global picture of the patterns that exist in the time series.

b. The division of multiple time series into training and testing parts where the training part is used to build the model, while the testing part is used to evaluate the model's performance on predicting unseen values.

c. The development of a VAR model which includes:

- Identification of the model structure by performing stationary tests, determining the optimum lags, and the Granger causality test.
- Formation of input-output pairs (predictor-target) based on the structure of the model.
- Parameter estimation (training) of the VAR model.
- Interpretation of VAR models and visualization of actual versus predicted values.

d. Evaluation of the performance of the VAR model. ARDL model development includes:

- Formation of input-output matrix based on the distribution of 7 lags from each time series as a predictor with the target being the current value of the time series.
- Selection of variables using the forward subset selection method.
- Parameter estimation (training) of the ARDL model. ARDL model interpretation and visualization of actual versus predicted values.
- ARDL model performance evaluation.

4. Result and Discussion

In this section, we will discuss the description of multiple time series data, distribution of training and testing sets, initialization of model structure, parameter estimation, evaluation of model performance in both training and testing sets for VAR and ARDL models.

4.1. Data Description

The data set used in this study was obtained from the information system on the availability and development of staple food prices in the province of East Java, Indonesia. Daily multiple time series data in the form of beef prices and chicken meat prices from March 20, 2020, to October 31, 2021. Fig. 1 presents the time series plots of the two data.

The modeling in this study uses a machine learning approach by dividing the dataset into two groups, namely the training part (2020 March 20 to 2021 August 31) and the testing part (September 1 to October 31, 2021).

The training set is used to build the model, while the testing part is used to select the best model that has the highest performance in the testing part. The daily data of the prices of both types of meat are presented in the form of a graph in Fig. 1 below.

Based on Fig. 1, information is obtained that the price of meat fluctuates from time to time. The price of chicken tends to fluctuate more than beef, although beef is relatively stable in certain periods there are quite large fluctuations.

The summary statistics of the two-time series data are presented in Table 1.

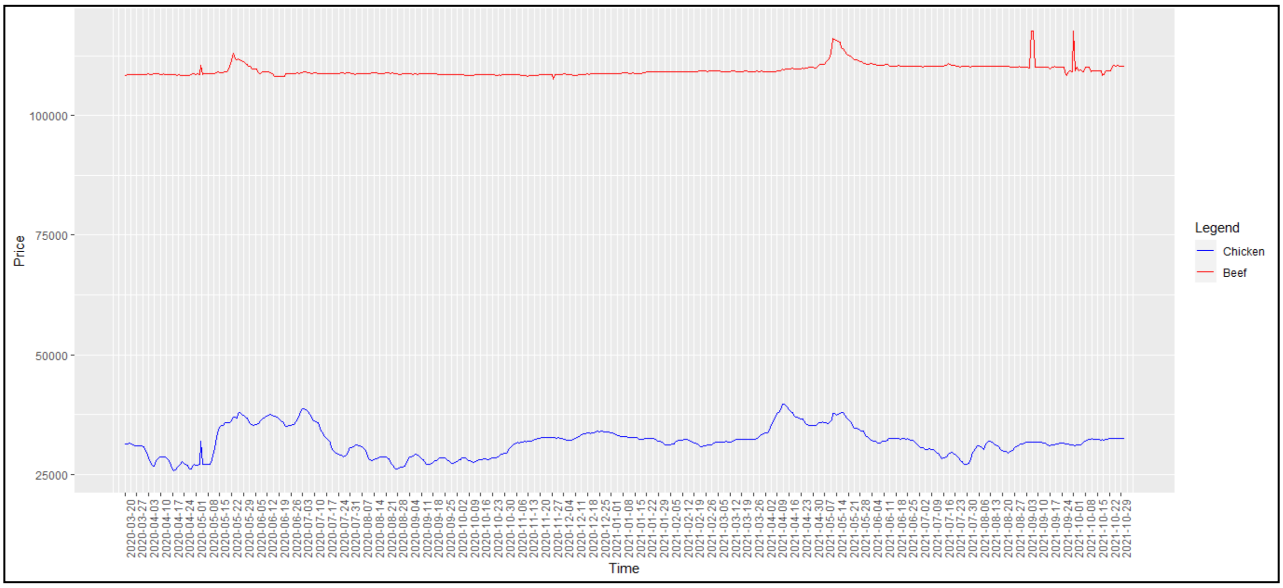


Fig. 1 The time series plot daily price of chicken and beef

Table 1 The summary statistic of the dataset

Price	Statistic					
	Min	**Q1**	**Median**	**Q3**	**Mean**	**Max**
Chicken	107670	108647	109092	110162	109435	117655
Beef	25729	29534	31811	33110	31814	39717

Based on table 1, the information is obtained that the price of beef in that period has an average price of IDR. 109435.00 with the lowest price of IDR. 107670.00 in the period of November 28, 2020. Although there had been fluctuations due to the covid19 pandemic, towards the end of the year beef prices began to stabilize again. Meanwhile, the highest price of beef is IDR. 117665.00 in the period of September 6, 2021. As many as 25% of the data has a price of less than IDR. 108647.00, as much as 50% of the data has a price of less than IDR. 109092.00 and as much as 25% of the data has a price of more than IDR.117655.00.

Meanwhile, the price of chicken meat during that period had an average price of IDR. 31814.00 with the lowest price being IDR. 25729.00 for the period of April 18, 2020. The first detection of the Covid19 case in early March 2020 caused people to be panic, so they tend to reduce activities outside the home. This causes the amount of demand for chicken meat prices to decline in the market. In addition, many business actors in the culinary field have lost a lot of customers. The highest price of chicken is IDR. 39717.00 at the period of April 12, 2021. This is because the inflation rate in the April 2021 period increased compared to the March 2021 period. The increase in prices for several food commodities in the month of Ramadan became the main factor in the occurrence of inflation in the April 2021 period. As many as 25% of the data had prices less than IDR. 29,534.00, 50% of the data have a price less than IDR. 31,811.00, and as much as 25% of the data has a price of more than IDR. 33,110.00.

4.2. VAR Modeling

In the VAR modeling, the data used is stationary time series data. So that the stationary data must be checked at a level. The result of the Augmented Dickey-Fuller (ADF) test concluded that at the level with an error rate of 5%, the beef price data in that period had a p-value (0.02828) < 0.05 (stationary) but the chicken meat price data in that period had p-value (0.4298)> 0.05 (not stationary). Because one of the data is not stationary, the first difference is carried out on the data and the ADF test is carried out again. The chicken data on changes after the first difference, the price of chicken meat in that period had a p-value (0.01) < 0.05 (stationary).

Before doing the VAR modeling, the optimum lag is selected to be included in the VAR model. Optimum lag selection can be done by selecting the lag that has the smallest AIC, HQ, SC, and FPE values. Table 2 shows the results of the optimum lag selection test:

Table 2 the statistic values of AIC, HQ, SC, FPE

Value	*Lag*						
	1	**2**	**3**	**4**	**5**	**6**	**7**
AIC	22.89	22.85	22.85	22.85	22.85	22.84*	22.85
HQ	22.91	22.88*	22.89	22.91	22.92	22.92	22.95
SC	22.94	22.93*	22.96	23.00	23.03	23.05	23.10
FPE	8,703E + 06	8,351E + 06	8,366E + 06	8,368E + 06	8,416E + 06	8,307E + 06*	8,388E + 06

Based on Table 2, it can be concluded that the optimum lags based on the values of AIC, HQ, SC, FPE are lag 2 and lag 6. Furthermore, these two lags are used to carry out the next process.

After the optimum lag is obtained, the next step is to do causality testing on each of these lags. The reciprocal relationship on the optimum lag can be evaluated by using the Granger causality test. The results of causality testing on the data differences on both chicken and beef prices with a 95% confidence level show that at lag 2 (p-value = 0.0007513 and p-value = 0.04383) there is sufficient evidence of a reciprocal relationship between changes in chicken and beef prices. While at lag 6 (p-value = 0.00816 and p-value = 0.1976), there is a one-way relationship, namely changes in chicken meat prices affect beef but not vice versa. So lag 2 is used to be included in the VAR modeling.

After identifying the VAR model structure, the next step is to estimate the parameters of the VAR model using the maximum likelihood method. The output of this stage is a VAR model that can be used to predict the price of chicken and beef in the future period. The following is the formula of the VAR model obtained:

$$\hat{s}_t = s_{t-1} + 0.20431(s_{t-1} - s_{t-2}) - 0.09639(a_{t-1} - \mathit{at{-}2{+}0.07094st{-}2{-}st{-}3{+}}\ 0.01402(a_{t-2} - a_{t-3})$$

$$\hat{s}_t = s_{t-1} + 0.20431(s_{t-1}) - 0.09639(a_{t-1}) - 0.13337(s_{t-2}) + 0.11041(a_{t-2}) - 0.07094(s_{t-3}) - 0.01402(a_{t-3}) \qquad (7)$$

$$\hat{a}_t = a_{t-1} - 0.23005(s_{t-1} - s_{t-2}) + 0.20672(a_{t-1} - a_{t-2}) - 0.01420(s_{t-2} - s_{t-3}) + 0.22233(a_{t-2} - a_{t-3})$$

$$\hat{a}_t = a_{t-1} - 0.23005(s_{t-1}) + 0.20672(a_{t-1}) + 0.21585(s_{t-2}) + 0.01561(a_{t-2}) + 0.01420(s_{t-3}) - 0.22233(a_{t-3}) \qquad (8)$$

where $\hat{s}_t$ is estimated value of beef prices in period t, $\hat{a}_t$ is estimated value of chicken meat price in period t, s_{t-k} is beef prices in period-(t-k), and a_{t-k} is chicken meat price in (t-k) period (t-k).

Based on equation 7, the information is obtained that the price of beef in this period is influenced by the price of beef from the previous period. The raising of 1 IDR in the price of beef in the previous period will increase the average price of beef for the current period by 0.20431 IDR and every one IDR increase in the price of chicken meat in the previous period will decrease the average price of beef in the current period by 0.09639 IDR. Each increase of 1 IDR of beef price in the previous two periods will reduce the current average price of beef by 0.13337 IDR and each increase of 1 IDR of chicken meat price in the two previous periods will increase the average beef price for the current period by 0.11041 IDR. Each increase in 1 IDR of beef price in the previous three periods will reduce the current average price of beef by 0.07094 IDR and each increase of 1 IDR of chicken meat price in the previous three periods will reduce the average price of beef for the current period by 0.01402 IDR. The interpretation of equation 8 is similar to the interpretation of equation 7.

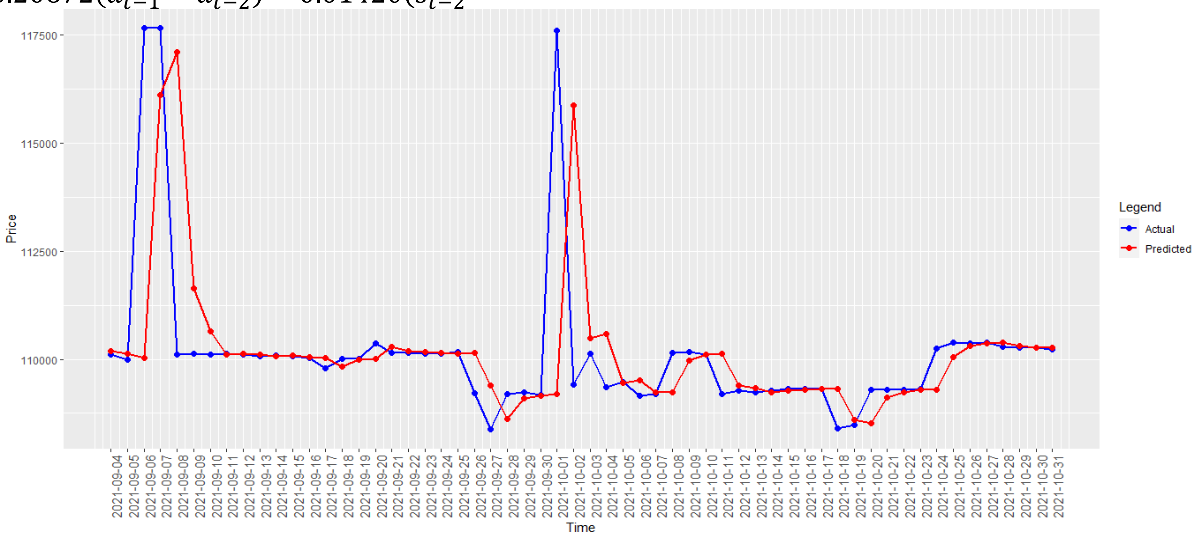


Fig. 2 Plot the actual versus predicted beef prices

Based on Fig. 2, the information is obtained that there is a large gap between the predicted values and the actual values. It can be seen from the graph that there is almost the same pattern in both the actual and predicted data, but it looks like a shift in the previous two lags. So if there is a fluctuation in the t period, it will only be detected in the t+2 period in the predicted value. This shows that the VAR model formed is less sensitive to shocks. Furthermore, the graph of the chicken meat price testing data is shown inn Fig. 3. The obtained information is that the predicted values are not close to the actual value. It can be seen from the graph that they have a different pattern between the actual and predicted values, and there are too sharp fluctuations in the predicted values. On another side, the actual data plot shows an up and down pattern but they are not too sharp. This means that an increase in the current actual value is followed by a slow increase of the previous period and vice versa.

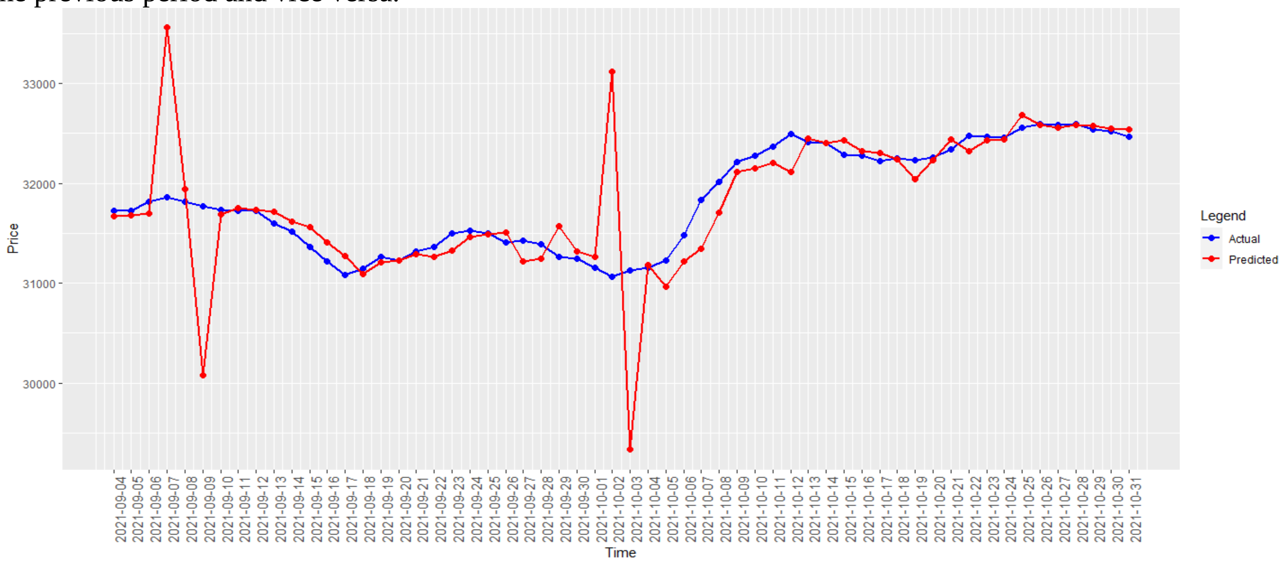
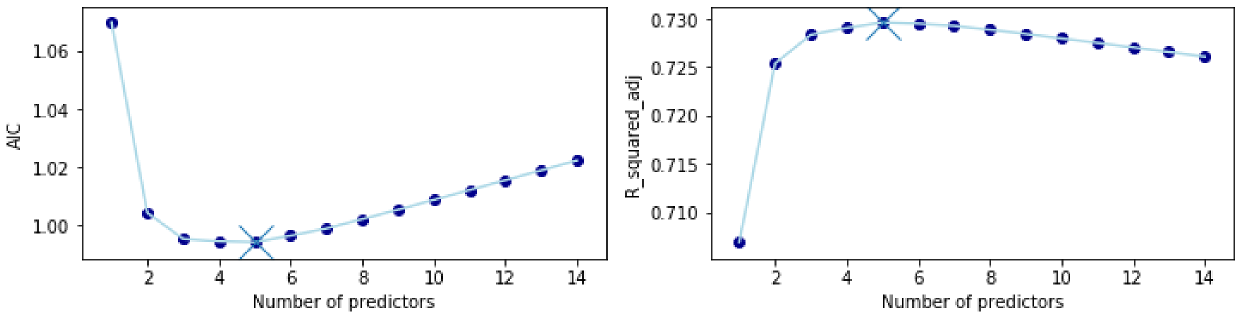


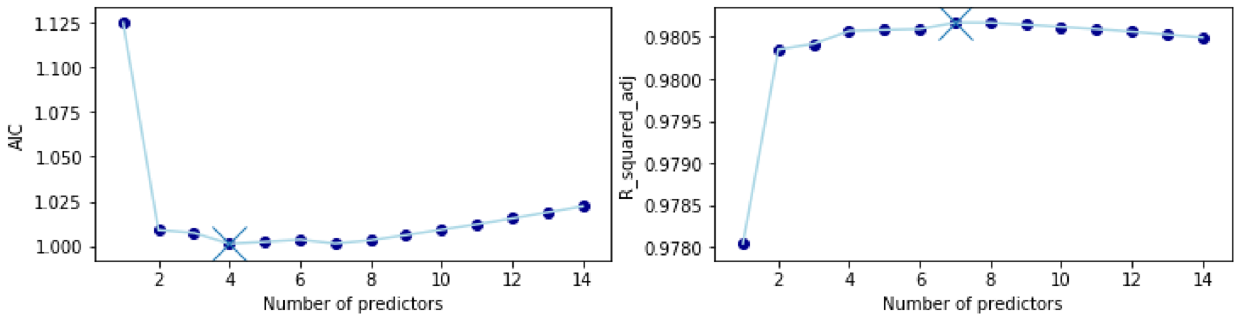
Fig. 3 Plot the actual versus predicted chicken meat prices

4.3. ARDL Modeling

Before estimating the parameters of the ARDL model on the price data for chicken and beef, the input-output pairs matrix is formed which includes the price of beef and chicken in period t as the target feature, while the prices of beef and chicken in the previous 7 periods are used as predictor variables. The selection of the optimum predictor variables is carried out separately for each input-output pairs matrix. The optimum input features are selected by using the forward subset method which is based on the criteria both of AIC and R square adjusted statistics. Fig. 4 presents the number of predictor variables that produce the optimum AIC and R square adjusted statistics.



a. The optimum beef price predictor features



b. The optimum chicken meat price predictor features

Fig. 4 The AIC and R square adjusted plot

Based on Fig. 4a, the number of beef price predictors selected is 5 features which are the beef price on the first lag (s_{t-1}), on the third lag (s_{t-3}), on the fifth lag (s_{t-5}), on the seventh lag (s_{t-7}), and the chicken price on the first lag (a_{t-1}). Based on Fig. 4b, the number of chicken meat price predictors selected is 4 features which are the chicken price on the first lag (a_{t-1}), on the fourth lag (a_{t-4}), on the fifth lag (a_{t-4}), and the seventh lag (a_{t-7}).

The ordinary least squares is carried out separately on each input-output pairs matrix with optimum features selected to obtain the estimated ARDL model parameters. The ARDL obtained is presented on equation 7 and equation 8 as the following:

$$\hat{s}_t = 1.07859(s_{t-1}) - 0.02823(s_{t-3}) - 0.12357(s_{t-5}) + 0.04217(s_{t-7}) + 0.00165(a_{t-1}) \quad (9)$$

$$\hat{a}_t = 1.14723(a_{t-1}) - 0.11081(a_{t-4}) - 0.13000(a_{t-5}) + 0.07771(a_{t-7}) \quad (10)$$

where $\hat{s}_t$ is the predicted value of the beef price and $\hat{a}_t$ is the estimated value of the chicken meat price at the period of t. Based on equation 9, the obtained information is the changes in beef prices for the current period are influenced by the beef prices in previous periods and the chicken prices in one previous period. Every IDR. 1 increase in the beef price in the previous period will increase the current average beef price by IDR. 1.07859, each increase in IDR. 1 of beef price in the previous three periods will decrease the average beef price for the current period by IDR. 0.02823. Each increase in IDR.1 of beef price in the previous five periods will reduce the average beef price for the current period by IDR. 0.12357, each increase in IDR. 1 of beef price in the previous seven periods will increase the current average beef price by IDR. 0.04217. Finally, every one IDR increase in the chicken meat price in the previous period will increase the current average beef price by IDR. 0.00165. The model interpretation of equation 10 is similar to the interpretation of the model in equation 9. The plot of the actual value versus the predicted beef price using the ARDL model is presented in Fig. 5.

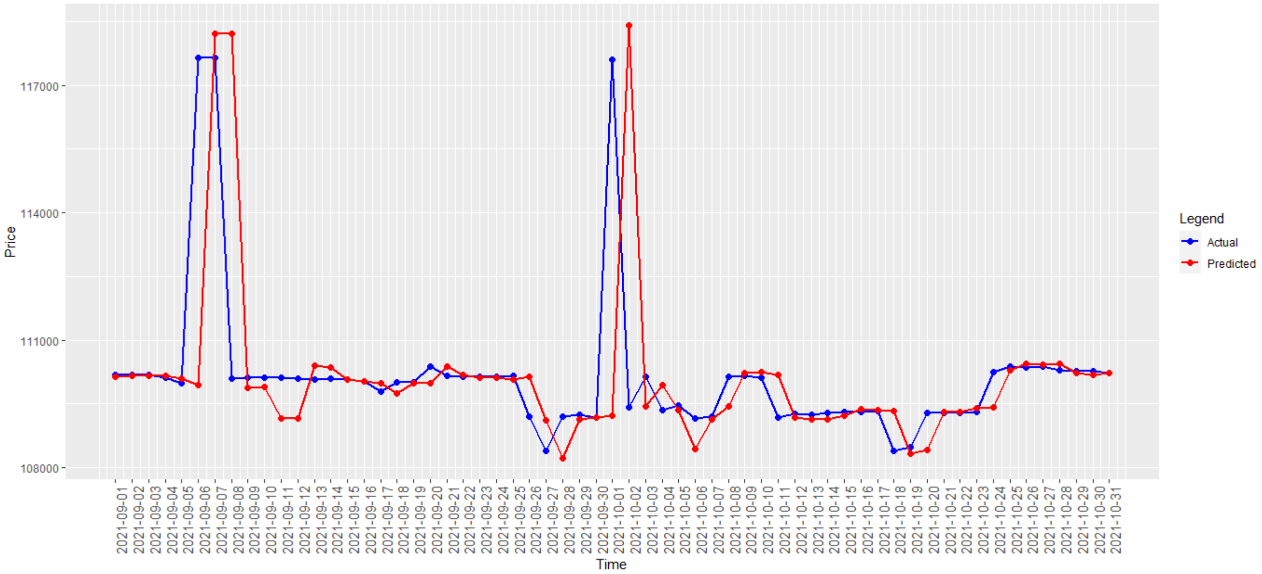


Fig. 5 The actual versus predicted values of the beef price by ARDL model

It can be seen that in general, the gap between the two values is quite narrow except for some price fluctuations where there is a wide gap, especially on the extreme value of beef price fluctuations at the first week of both 2021 September and October.

Fig. 6 presents a plot of the actual versus the predicted values of the price of chicken meat using the model in equation (10). The pattern of the actual values of chicken meat prices can be well followed by predictive values with a narrow gap between them. The implication of this is that the model in equation (10) can explain well the existing patterns in the price of chicken meat.

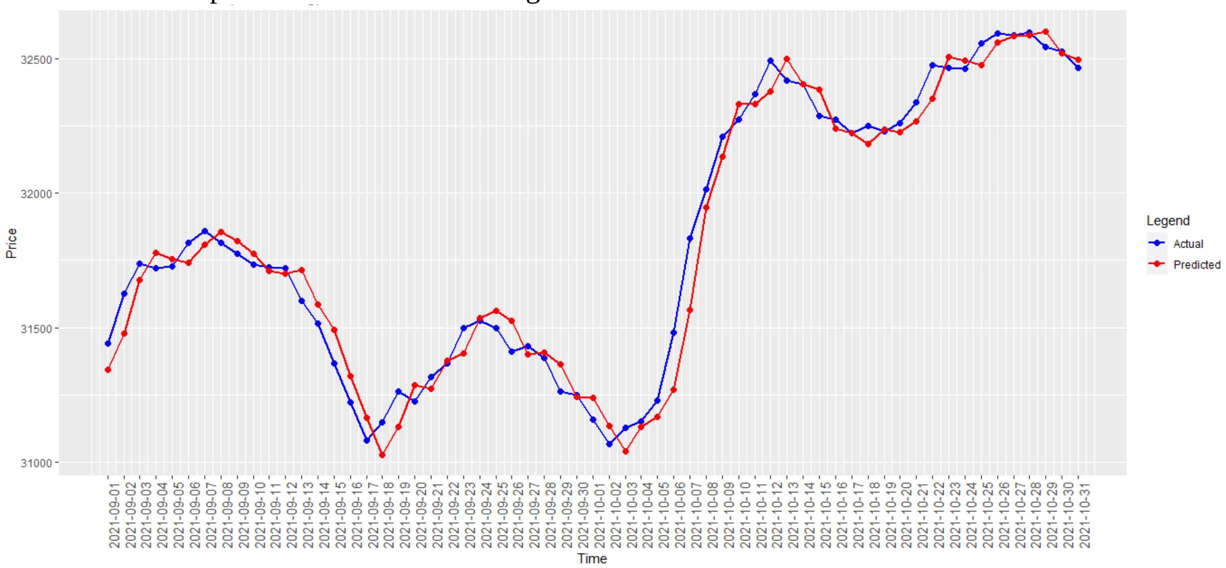


Fig. 6 The actual versus predicted values of the chicken meat price by ARDL model

4.4. The Comparison between VAR and ARDL Model Performance

In addition to visual evaluation through graphs, the model performance of both models is also evaluated by using Root Mean Square Error (MSE), Mean Absolute Error (MAE), and the value of the correlation squared (R square). Table 3 presents the VAR model performance in both training and testing sets.

Table 3 The performance measures of the VAR model

Criteria	Beef price		Chicken meat price	
	Training data	Testing data	Training data	Testing data
RMSE	236.12	2006.33	446.44	497.53
MAE	102.03	777.56	218.98	226.49
R square	0.95986	0.09377	0.9806	0.4879

Based on Table 3, it can be concluded that in the training set, all of the criteria including the RMSE, MAE, and R square show a fairly good value, namely small MAE and RMSE values and high R square values. But the beef price testing set, the value of the R square is very low (9.5%) and the MAE and RMSE values are very higher than the training set (more 700 % than the training set). While the chicken meat pride testing set has all of the criteria that are better than the beef price testing set which indicates that the model in equation (8) is better than the model in equation (7). This information leads to the fact that the model in equation (7) and equation (8) in forecasting beef and chicken meat prices suffers an over-fitting or overtraining.

Table 4 The performance measures of the ARDL model

Criteria	Beef price		Chicken meat price	
	Training data	Testing data	Training data	Testing data
RMSE	231.93	2168.68	443.72	81.56
MAE	98.90	804.29	214.92	64.51
R square	0.9592	0.1034	0.9807	0.9732

The performance of both ARDL models in this study is better than the VAR model which is indicated by its much higher R square value, but if there are extreme fluctuations in price values, the VAR model performs slightly better (has a smaller RMSE and MAE). This is because, in the process of estimating the parameters of the VAR model, it also considers the cross-correlation between time series. The R square value of the ARDL model in the testing set is higher, indicating that the fitting between the actual versus predicted values in the ARDL model testing set is better than the VAR model. The R square value of the ARDL model for chicken meat prices is very high if it refers to equation (4) this ARDL model does not involve a lag distribution of beef prices, so the model in equation (4) is an autoregressive (AR) model which involves 4 lags of the chicken meat price.

Initialization of the model structure has a very important role in producing a good model. Theoretically, VAR modeling has a more complex model initialization stage than ARDL modeling. In this study, forward subset selection which is a feature selection method in the wrapped method group is theoretically simple but computationally expensive.

5. Conclusion

Some notices that should be underlined in the modeling of multiple time series using the VAR and ARDL models with a machine learning approach include they have different model structures and different performance significantly in the case of the multiple time series studied in this work. The VAR model has the same predictors of the lag inputs on both of the beef and chicken meat prices model, but the ARDL model of both beef and chicken meat prices has different predictors of the lag inputs. The different model structure between VAR and ARDL modeling is caused by the different approaches in the stage of identifying model structure which has a significant effect on their performance.

The result from the VAR model has the predictor lags i.e. lag 1, lag 2 lag 3 of both beef and chicken price. The ARDL model of beef price has the predictor lags i.e. lag 1, lag 3, lag 5, lag 7 of the beef price, and lag 1 of the chicken meat price. The ARDL model of chicken meat price has the predictor lags i.e. lag 1, lag 4, lag 5, lag 7 of the chicken meat price. The RMSE values of the beef price VAR model are IDR 236.12 and IDR 2006.33 respectively in the training and testing part. The RMSE values of the chicken meat price VAR model are IDR 446.44 and IDR 497.53 respectively on the training and testing part. The beef price VAR model suffers the overfitting problem which is supported by its R square value of 98.06% and 48.79% respectively on the training and testing part of the chicken meat price model.

The RMSE values of the beef price ARDL model are IDR 231.93 and IDR 2168.68 respectively in the training and testing part. The RMSE values of the chicken meat price VAR model are IDR 443.72 and IDR 81.56 respectively in the training and testing part. The ARDL model has better performance than the VAR model which is shown by its R square value of 98.07% and 97.32% respectively on the training and testing part of the chicken meat price model.

We recommend that the application of VAR on multiple time series requires the condition that both time series have bidirectional causality. Because the models that make up the VAR must have the same model structure. On the other hand, ARDL modeling does not require 2-way causality of both time series modeled, because ARDL modeling can be done by building a single model. While the VAR model must involve the number of constituent models equal to the number of time series being modeled.

The forward subset variable selection method proposed by this study to identify the structure of the ARDL model can produce lag variables as input variables (predictors) of the ARDL model that performs very satisfactorily. Visualization of the values of AIC and R square adjusted for all the number of possible predictor variables for the ARDL model which can provide a comprehensive point of view to the performance of the ARDL model. On the other hand, the identification of the VAR model structure using the conventional approach method in time series modeling is more intuitive and cannot guarantee the satisfactory performance of the VAR model. In the next research, an idea to be able to improve the performance of the VAR model and at the same time be able to overcome the problem of overfitting is VAR modeling by adding regularization of L2 norm or L1 norm to form a Ridge VAR or Lasso VAR model. Likewise, the performance of the ARDL model may also be improved by constructing a Ridge ARDL or Lasso ARDL model.

References

[1] NUGROHO W. H., HANDOYO S., and AKRI Y. J. An Influence of Measurement Scale of Predictor Variable on Logistic Regression Modeling and Learning Vector Quantization Modeling for Object Classification. *International Journal of Electrical and Computer Engineering*, 2018, 8(1): 333-343. http://dx.doi.org/10.11591/ijece.v8i1.pp333-343

[2] HANDOYO S., & KUSDARWATI H. Implementation of Fuzzy Inference System for Classification of Dengue Fever on the villages in Malang. *Basic Science International Conferences*, 2019, 546(5): 052026. https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052026/pdf

[3] HANDOYO S., CHEN Y. P., IRIANTO G., and WIDODO A. The Varying Threshold Values of Logistic Regression and Linear Discriminant for Classifying Fraudulent Firm. *Mathematics and Statistics*, 2021, 9(2): 135-143. http://dx.doi.org/10.13189/ms.2021.090207

[4] UTAMI H. N., CANDRA, and HANDOYO S. The Effect of Self Efficacy and Hope on Occupational Health Behavior in East Java of Indonesia. *International Journal of Scientific & Technology Research*, 2020, 9(2): 3571-3575. http://www.ijstr.org/final-print/feb2020/The-Effect-Of-Self-Efficacy-And-Hope-On-Occupational-Health-Behavior-In-East-Java-Of-Indonesia.pdf

[5] KUSDARWATI H., and HANDOYO S. Modeling Treshold Liner in Transfer Function to Overcome Non Normality of the Errors. *Basic Science International Conferences*, 2019, 546(5): 052039. https://www.researchgate.net/publication/332094380_Modeling_Treshold_Liner_in_Transfer_Function_to_Overcome_Non_Normality_of_the_errors

[6] HANDOYO S., MARJI, PURWANTO I. N., and JIE F. The Fuzzy Inference System with Rule Bases Generated by using the Fuzzy C-Means to Predict Regional Minimum Wage in Indonesia. *International Journal of Operations and Quantitative Management*, 2018, 24(4): 277-292. https://www.researchgate.net/publication/331530844_The_Fuzzy_Inference_System_with_Rule_Bases_Generated_by_using_the_Fuzzy_C-Means_to_Predict_Regional_Minimum_Wage_in_Indonesia

[7] KUSDARWATI H., & HANDOYO S. System for Prediction of Non Stationary Time Series based on the Wavelet Radial Bases Function Neural Network Model. *International Journal Electrical & Computer Engineering*, 2018, 8(4): 2327-2337. http://doi.org/10.11591/ijece.v8i4.pp2327-2337

[8] PUROHIT S. K., PANIGRAHI S., SETHY P. K., and BEHERA S. K. Time series forecasting of price of agricultural products using hybrid methods. *Applied Artificial Intelligence*, 2021: 1-19. https://doi.org/10.1080/08839514.2021.1981659

[9] PRAVILOVIC S., BILANCIA M., APPICE A., and MALERBA D. Using multiple time series analysis for geosensor data forecasting. *Information Sciences,* 2017, 380: 31-52. http://dx.doi.org/10.1016/j.ins.2016.11.001

[10] SHIH S. Y., SUN F. K., and LEE H. Y. Temporal pattern attention for multivariate time series forecasting. *Machine Learning,* 2018, 108(8): 1421-1441. https://doi.org/10.48550/arXiv.1809.04206

[11] HANDOYO S., & CHEN Y. P. The Developing of Fuzzy System for Multiple Time Series Forecasting with Generated Rule Bases and Optimized Consequence Part. *International Journal of Engineering Trends and Technology*, 2020, 68(12): 118-122. http://dx.doi.org/10.14445/22315381/IJETT-V68I12P220

[12] MARICA V. G., and HOROBET A. Conditional Granger Causality and Genetic Algorithms in VAR Model Selection. *Symmetry*, 2019, 11(8): 1004. http://dx.doi.org/10.3390/sym11081004

[13] CHEN Y., & LIU D. Government spending shocks and the real exchange rate in China: Evidence from a sign-restricted VAR model. *Economic Modelling*, 2018, 68: 543-554. http://dx.doi.org/10.1016/j.econmod.2017.03.027

[14] CHANDIO A. A., JIANG Y., RAUF A., AHMAD F., AMIN W., and SHEHZAD K. Assessment of formal credit and climate change impact on agricultural production in Pakistan: a time series ARDL modeling approach. *Sustainability*, 2020, 12(13): 5241. https://doi.org/10.3390/su12135241

[15] SOHAIL M. T., ULLAH S., MAJEED M. T., and USMAN A. Pakistan management of green transportation and environmental pollution: a nonlinear ARDL analysis. *Environmental Science and Pollution Research*, 2021, 28: 29046–29055. https://doi.org/10.1007/s11356-021-12654-x

[16] IRCIO J., LOJO A., MORI U., and LOZANO J. A. Mutual information based feature subset selection in multivariate time series classification. *Pattern Recognition*, 2020, 108: 107525. https://doi.org/10.1016/j.patcog.2020.107525

[17] ZHU X., DU W., GENG G., and XUE B. Research on pork price prediction based on multi-dimensional feature analysis and machine learning. Proceedings of the 2021 ACM International Conference on Intelligent Computing and its Emerging Applications, 2021, pp. 111-116. https://doi.org/10.1145/3491396.3506521

[18] ZHANG H., WANG J., and MARTIN W. Factors affecting households' meat purchase and future meat consumption changes in China: a demand system approach. *Journal of Ethnic Foods*, 2018, 5(1): 24-32.

https://doi.org/10.1016/j.jef.2017.12.004
[19] LUSK J. L., TONSOR G. T., and SCHULZ L. L. Beef and pork marketing margins and price spreads during COVID 19. *Applied Economic Perspectives and Policy*, 2021, 43(1): 4-23. https://doi.org/10.1002/aepp.13101
[20] AGUIRRE A., & AGUIRRE L. A. Time series analysis of monthly beef cattle prices with nonlinear autoregressive models. *Applied Economics*, 2000, 32(3): 265-275. https://doi.org/10.1080/000368400322697
[21] ZENG B., LI S., MENG W., and ZHANG D. An improved gray prediction model for China's beef consumption forecasting. *PloS One*, 2019, 14(9): 0221333. https://doi.org/10.1371/journal.pone.0221333
[22] ALDERINY M. M., ALRWIS K. N., AHMED S. B., and ALDAWDAHI N. M. Forecasting Saudi Arabia's production and imports of broiler meat chickens and its effect on expected self-sufficiency ratio. *Journal of the Saudi Society of Agricultural Sciences*, 2020, 19(4): 306-312. https://doi.org/10.1016/j.jssas.2019.09.001
[23] PAPARODITIS E., & POLITIS D. N. The asymptotic size and power of the augmented Dickey–Fuller test for a unit root. *Econometric Reviews*, 2018, 37(9): 955-973. https://doi.org/10.1080/00927872.2016.1178887
[24] SHRESTHA M. B., & BHATTA G. R. Selecting appropriate methodological framework for time series data analysis. *The Journal of Finance and Data Science*, 2018, 4(2): 71-89. https://doi.org/10.1016/j.jfds.2017.11.001
[25] PERVUKHINA E., EMMENEGGER J. F., GOLIKOVA V., and OSIPOV K. An optimization technique based on a vector autoregression model with state space representation: application to Ukrainian cargo transport data. *Optimization*, 2014, 63(1): 93-108. https://doi.org/10.1080/02331934.2013.836648
[26] SARKER B., & KHAN F. Nexus between foreign direct investment and economic growth in Bangladesh: an augmented autoregressive distributed lag bounds testing approach. *Financial Innovation*, 2020, 6(1): 1-18. https://ideas.repec.org/a/spr/fininn/v6y2020i1d10.1186_s40854-019-0164-y.html
[27] CEESAY E. K., FRANCIS P. C., JAWNEH S., NJIE M., BELFORD C., and FANNEH M. M. Climate change, growth in agriculture value-added, food availability and economic growth nexus in the Gambia: a Granger causality and ARDL modeling approach. *SN Business & Economics*, 2021, 1(7): 1-31. https://econpapers.repec.org/scripts/redir.pf?u=https%3A%2F%2Fdoi.org%2F10.1007%252Fs43546-021-00100-6;h=repec:spr:snbeco:v:1:y:2021:i:7:d:10.1007_s43546-021-00100-6
[28] HASTIE T., TIBSHIRANI R., and TIBSHIRANI R. Best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons. *Statistical Science*, 2020, 35(4): 579-592. https://doi.org/10.1214/19-STS733
[29] SMITH G. Step away from stepwise. *Journal of Big Data*, 2018, 5(1): 1-12. https://doi.org/10.1186/s40537-018-0143-6
[30] HANDOYO S., & MARJI. The Fuzzy Inference System with Least Square Optimization for Time Series Forecasting. *Indonesian Journal of Electrical Engineering and Computer Science*, 2018, 7(3): 1015-1026. http://doi.org/10.11591/ijeecs.v11.i3.pp1015-1026

参考文:

[1] NUGROHO W. H., HANDOYO S., 和 AKRI Y. J. 预测变量的测量尺度对对象分类的逻辑回归建模和学习向量量化建模的影响。国际电气与计算机工程杂志, 2018, 8(1): 333-343. http://dx.doi.org/10.11591/ijece.v8i1.pp333-343
[2] HANDOYO S., 和 KUSDARWATI H. 登革热分类模糊推理系统在玛琅村庄的实施。基础科学国际会议, 2019, 546(5): 052026. https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052026/pdf
[3] HANDOYO S., CHEN Y. P., IRIANTO G., 和 WIDODO A. 用于分类欺诈公司的逻辑回归和线性判别的可变阈值。数学和统计, 2021, 9(2): 135-143. http://dx.doi.org/10.13189/ms.2021.090207
[4] UTAMI H. N., CANDRA, 和 HANDOYO S. 自我效能感和希望对印度尼西亚东爪哇职业健康行为的影响。国际科技研究杂志, 2020, 9(2): 3571-3575. http://www.ijstr.org/final-print/feb2020/The-Effect-Of-Self-Efficacy-And-Hope-On-Occupational-Health-Behavior-In-East-Java-Of-Indonesia.pdf
[5] KUSDARWATI H., 和 HANDOYO S. 在传递函数中建模阈值线性以克服错误的非正态性。基础科学国际会议, 2019, 546(5): 052039. https://www.researchgate.net/publication/332094380_Modeling_Treshold_Liner_in_Transfer_Function_to_Overcome_Non_Normality_of_the_errors
[6] HANDOYO S., MARJI, PURWANTO I. N., 和 JIE F. 使用模糊 C 均值生成的具有规则库的模糊推理系统来预测印度尼西亚的区域最低工资。国际运营与量化管理杂志, 2018, 24(4): 277-292. https://www.researchgate.net/publication/331530844_The_Fuzzy_Inference_System_with_Rule_Bases_Generated_by_using_the_Fuzzy_C-Means_to_Predict_Regional_Minimum_Wage_in_Indonesia
[7] KUSDARWATI H., 和 HANDOYO S. 基于小波径向基函数神经网络模型的非平稳时间序列预测系统。国际期刊电气与计算机工程, 2018, 8(4): 2327-2337. http://doi.org/10.11591/ijece.v8i4.pp2327-2337
[8] PUROHIT S. K., PANIGRAHI S., SETHY P. K., 和 BEHERA S. K. 使用混合方法的农产品价格时间序列预测。应用人工智能, 2021: 1-19. https://doi.org/10.1080/08839514.2021.1981659
[9] PRAVILOVIC S., BILANCIA M., APPICE A., 和 MALERBA D. 使用多时间序列分析进行地理传感器数据预测。信息科学, 2017, 380: 31-52. http://dx.doi.org/10.1016/j.ins.2016.11.001
[10] SHIH S. Y., SUN F. K., 和 LEE H. Y. 多变量时间序列预测的时间模式注意。机器学习, 2018, 108(8): 1421-1441. https://doi.org/10.48550/arXiv.1809.04206
[11] HANDOYO S., 和 CHEN Y. P. 具有生成规则库和优化结果部分的多时间序列预测模糊系统的开发。国际工程趋势与技术杂志, 2020, 68(12): 118-122. http://dx.doi.org/10.14445/22315381/IJETT-

V68I12P220
[12] MARICA V. G., 和 HOROBET A. 向量自回归模型选择中的条件格兰杰因果关系和遗传算法。对称, 2019, 11(8): 1004. http://dx.doi.org/10.3390/sym11081004
[13] CHEN Y., 和 LIU D. 中国政府支出冲击和实际汇率：来自符号限制向量自回归模型的证据。经济建模, 2018, 68: 543-554. http://dx.doi.org/10.1016/j.econmod.2017.03.027
[14] CHANDIO A. A., JIANG Y., RAUF A., AHMAD F., AMIN W., 和 SHEHZAD K. 评估正式信贷和气候变化对巴基斯坦农业生产的影响：时间序列自回归分布式滞后建模方法。可持续性, 2020, 12(13): 5241. https://doi.org/10.3390/su12135241
[15] SOHAIL M. T., ULLAH S., MAJEED M. T., 和 USMAN A. 巴基斯坦绿色交通和环境污染管理：非线性自回归分布滞后分析。环境科学与污染研究, 2021, 28: 29046–29055. https://doi.org/10.1007/s11356-021-12654-x
[16] IRCIO J., LOJO A., MORI U., 和 LOZANO J. A. 多元时间序列分类中基于互信息的特征子集选择。模式识别, 2020, 108: 107525. https://doi.org/10.1016/j.patcog.2020.107525
[17] ZHU X., DU W., GENG G., 和 XUE B基于多维特征分析和机器学习的猪肉价格预测研究2021年计算机协会智能计算及其新兴应用国际会议论文集, 2021, 页。111-116. https://doi.org/10.1145/3491396.3506521
[18] ZHANG H., WANG J., 和 MARTIN W. 影响中国家庭肉类购买和未来肉类消费变化的因素：一种需求系统方法。民族食品杂志, 2018, 5(1): 24-32. https://doi.org/10.1016/j.jef.2017.12.004
[19] LUSK J. L., TONSOR G. T., 和 SCHULZ L. L. 新冠肺炎期间牛肉和猪肉的营销利润和价差。应用经济观点和政策, 2021, 43(1): 4-23. https://doi.org/10.1002/aepp.13101
[20] AGUIRRE A., 和 AGUIRRE L. A. 使用非线性自回归模型对每月肉牛价格进行时间序列分析。应用经济学, 2000, 32(3): 265-275. https://doi.org/10.1080/000368400322697
[21] ZENG B., LI S., MENG W., 和 ZHANG D. 中国牛肉消费预测的改进灰色预测模型.科学一号公共图书馆, 2019, 14(9): 0221333. https://doi.org/10.1371/journal.pone.0221333
[22] ALDERINY M. M., ALRWIS K. N., AHMED S. B., 和 ALDAWDAHI N. M. 预测沙特阿拉伯肉鸡的生产和进口及其对预期自给率的影响。沙特农业科学学会杂志, 2020, 19(4): 306-312. https://doi.org/10.1016/j.jssas.2019.09.001
[23] PAPARODITIS E., 和 POLITIS D. N. 单位根的增广迪基-富勒检验的渐近大小和幂。计量经济学评论, 2018, 37(9): 955-973. https://doi.org/10.1080/00927872.2016.1178887
[24] SHRESTHA M. B., 和 BHATTA G. R. 为时间序列数据分析选择适当的方法框架。金融与数据科学杂志, 2018, 4(2): 71-89. https://doi.org/10.1016/j.jfds.2017.11.001
[25] PERVUKHINA E., EMMENEGGER J. F., GOLIKOVA V., 和 OSIPOV K. 基于具有状态空间表示的向量自回归模型的优化技术：在乌克兰货物运输数据中的应用。优化, 2014, 63(1): 93-108. https://doi.org/10.1080/02331934.2013.836648
[26] SARKER B., 和 KHAN F. 孟加拉国外国直接投资与经济增长之间的关系：一种增强的自回归分布式滞后边界测试方法。金融创新, 2020, 6(1): 1-18. https://ideas.repec.org/a/spr/fininn/v6y2020i1d10.1186_s40854-019-0164-y.html
[27] CEESAY E. K., FRANCIS P. C., JAWNEH S., NJIE M., BELFORD C., 和 FANNEH M. M. 冈比亚的气候变化、农业附加值增长、粮食供应和经济增长关系：格兰杰因果关系和自回归分布式滞后建模方法。施普林格自然商业与经济学, 2021, 1(7): 1-31. https://econpapers.repec.org/scripts/redir.pf?u=https%3A%2F%2Fdoi.org%2F10.1007%252Fs43546-021-00100-6;h=repec:spr:snbeco:v:1:y:2021:i:7:d:10.1007_s43546-021-00100-6
[28] HASTIE T., TIBSHIRANI R., 和 TIBSHIRANI R. 最佳子集，向前逐步还是套索？基于广泛比较的分析和建议。统计科学, 2020, 35(4): 579-592. https://doi.org/10.1214/19-STS733
[29] SMITH G. 一步一步远离。大数据杂志, 2018, 5(1): 1-12. https://doi.org/10.1186/s40537-018-0143-6
[30] HANDOYO S., 和 MARJI. 用于时间序列预测的具有最小二乘优化的模糊推理系统。印度尼西亚电气工程与计算机科学杂志, 2018, 7(3): 1015-1026. http://doi.org/10.11591/ijeecs.v11.i3.pp1015-1026