Songklanakarin J. Sci. Technol.
46 (4), 377–385, Jul. – Aug. 2024

http://www.sjst.psu.ac.th

Original Article

Forecasting food prices in Central Java with the hybrid ARIMAX calendar variation model

Vebryana Wahyuningsih*, and Putriaji Hendikawati

Department of Mathematics, Faculty of Mathematics and Natural Science, Semarang State University, Semarang, 50229 Indonesia

Received: 23 November 2023; Revised: 1 April 2024; Accepted: 27 May 2024

Abstract

Variations in time series data that occur due to the effects of calendar factors such as holidays, seasons, or differences in the number of days in a month are referred to as calendar variations. Time series problems are not always linear or nonlinear. Instead, they sometimes contain both (linear and nonlinear) at once hence hybrid models can be used to model time series problems. Research was conducted on food prices in Central Java that are affected by calendar variations, namely beef and chicken egg price data. The performance of these methods is compared using RMSE and MAPE accuracy measures. The hybrid ARIMAX-NN and Reg-ARIMA methods produce the smallest RMSE and MAPE values compared to other methods, respectively.

Keywords: time series data, calendar variation, hybrid method, foodstuffs

1. Introduction

Calendar variation in time series data is attributed to fluctuations caused by calendar factors like holidays, seasons, and variations in the number of days per month. Economic and commercial monthly data are often affected by two calendar effects: (1) the trading day effect, influenced by the distribution of working days in a month, and (2) the holiday effect, resulting from changes in the dates of festivals and public holidays. These variations can significantly impact on the observed time series, depending on the specific month and year (Bell & Hillmer, 1983; Gao & Kling, 2005; Hillmer, 1982; Liu, 1986; Mills & Coutts, 1995; Nisa, Tarno, & Rusgiyono, 2020; Suhartono, 2006).

Time series problems can exhibit both linear and nonlinear characteristics simultaneously, making it challenge for a single model to capture all attributes. To address this, hybrid models, combining both linear and nonlinear elements, are commonly employed in various studies to enhance the effectiveness of time series analysis (Pavlenko & Matvejevs, 2023; Terui & Dijk, 2002). There are several hybrid methods that are usually used for forecasting time series data with calendar variations. Some previous studies using hybrid models include Jesica, Ispriyanti, & Tarno (2019) using the Reg-ARIMA model to forecast the number of tourists visiting tourist attractions in Central Java, Nisa *et al.* (2020) using the Reg-ARIMA model to forecast red chili prices with calendar variations, Kusumaningrum, Purnamasari, & Siringoringo (2023) using the ARIMAX-NN model to forecast total non-cash payment transactions, Sihombing, Martha, & Huda (2022) using the ARIMA-SVR model to model JCI data, and Chow, (2021), which forecasted the short-term maximum demand of the electricity network with ARIMAX-SVR.

In Central Java, a densely populated province in Indonesia, the prices of essential food items like beef and chicken eggs from May 2017 to August 2023, as recorded in the National Strategic Food Price Information Center, have consistently risen. This upward trend can potentially impact on individuals, the economy, and the overall society. To model the price data, it is essential to consider fluctuations influenced by calendar variations, including religious holidays like Eid al-Fitr, Imlek Day, and Christmas. Implementing a hybrid method for handling calendar variations is recommended to address these effects in the time series data. It is important to forecast food commodity prices because it helps producers plan production, assists policymakers in determining appropriate economic policies, and ensures food security (Pavlenko & Matvejevs, 2023).

*Corresponding author
Email address: vebryanaw@students.unnes.ac.id

This study aims to identify and compare the performance of beef and chicken egg price data in predicticy ARIMAX, ARIMAX-NN, ARIMAX-SVR and Reg-ARIMA. The reason these methods are used is that they can improve forecast accuracy by adding independent variables as exogenous or dummy variables (Anggraeni, Vinarti, & Kurniawati, 2015; Prastuti, Suhartono, & Salehah, 2018; Rahmayani, 2013). The machine learning model automatically updates itself when new data is input (Chow, 2021). The hybrid models aim to combine the strengths of different models so that the risk of model misspecification can be reduced. In addition, the method belongs to the soft computing calculation method where this calculation can more accurately model nonlinear relationships and avoid making many assumptions about the data set (Hendikawati, Subanar, Abdurakhman, & Tarno, 2021). The data used in this study are food price data for beef and chicken eggs in Central Java from May 2017 to August 2023 obtained from the National Strategic Food Price Information Center. This scientific work is organized as follows: In section 2, the methods and materials used in the research are described. In section 3, case study results obtained using forecasting methods are presented to compare the performance of the methods. In section 4, conclusions are drawn and an outline of the research results is provided to be used as suggestions for further research.

2. Materials and Methods

2.1 Data used

The study utilized monthly data spanning from May 2017 to August 2023, focusing on beef and chicken egg prices sourced from *www.bi.go.id/hargapangan*. The selected timeframe anticipates significant price increases. While beef price data showed non-stationarity, chicken egg prices remained stationary, with no discontinuities in the time series. The study incorporated exogenous variables reflecting holiday effects, such as Eid al-Fitr, Chinese New Year, and Christmas. The 76 monthly data of beef and chicken egg prices are divided into two parts, 80% as training data, and 20% as testing data.

2.2 Exogenous variables

Exogenous variables, impacting other variables but unaffected by them, are used in time series data with calendar variations, typically represented by dummy variables (Riska & Hasi, 2019). The calendar variation used in this study is the type of holiday effect, namely Eid al-Fitr, Chinese New Year, and Christmas. According to the dummy variable is 1 for one month before the holiday, the month during the holiday, and one month after the holiday. Furthermore, the dummy variable is 0 for other months (Intan, Zukhronah, & Wibowo, 2018; Kusumaningrum *et al.*, 2023; Riska & Hasi, 2019). The dummy variable is also a regressor variable in the regression model.

2.3 ARIMAX model

The development of the ARIMA model with the addition of exogenous variables as explanatory variables is referred to as ARIMAX. ARIMA and ARIMAX modeling involves checking the stationarity of time series data using the augmented Dickey-Fuller (ADF) test, as well as analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ARIMAX model emphasizes independent variable patterns for accurate forecasting (Lestari, 2023; Rahmayani, 2013). The ARIMAX model equation includes a dummy variable input.

$$Y_t = \beta_1 A_{t-1} + \beta_2 A_t + \beta_3 A_{t+1} + \gamma_1 B_{t-1} + \gamma_2 B_t \gamma_3 B_{t+1} + \delta_1 C_{t-1} + \delta_2 C_t + \delta_3 C_{t+1} + \frac{\theta_q(B)}{\phi_p(B)} a_t$$

with p is ordo autoregressive (AR), t is the t-th observation data, $t = 1,2,3,\ldots$, q is ordo moving average (MA), $\phi_p(B)$ is $1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\theta_q(B)$ is $1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, ϕ is autoregressive coefficient (AR), θ is moving average coefficient (MA), a_t is t-th error, $\beta_1, \beta_2, \beta_3$ are coefficient one month before, during, one month after Chinese New Year. A_{t-1}, A_t, A_{t+1} are dummy variable one month before, on, one month after Chinese New Year. $\gamma_1, \gamma_2, \gamma_3$ are coefficient one month before, on, one month after Eid al-Fitr. B_{t-1}, B_t, B_{t+1} are dummy variable one month before, on, one month after Eid al-Fitr. $\delta_1, \delta_2, \delta_3$ are coefficient one month before, at the time of, one month after Christmas. C_{t-1}, C_t, C_{t+1} are dummy variable one month before, on, one month after Christmas

2.4 Hybrid ARIMAX-NN model

Neural network (NN) is an information processing method based on modeling biological neural networks, particularly the human brain (Zhang, 2004). A typical NN comprises three layers: input, hidden, and output. A popular model is the feed forward neural network, represented by following specific equation.

$$M_t = f^0 \sum_{j=1}^{q} \left\{ W_j^0 . f_j^{\boldsymbol{h}} \left(\sum_{i=1}^{p} W_{ji}^{\boldsymbol{h}} . N_{t-1} + b_j^{\boldsymbol{h}} \right) + b^0 \right\}$$

with M_t is output variable, f^0 is activation function at the neuron in the output layer, q is number of hidden layers, W_j^0 is the weight of the j-th neuron in the hidden layer that goes to the output layer, $f_j^{\boldsymbol{h}}$ is j-th neuron activation function, p is number of input layers, $W_{ji}^{\boldsymbol{h}}$ is the weight of the I-th input that goes to the hidden layer, N_{t-1} is input variable, $b_j^{\boldsymbol{h}}$ is bias of neurons in the hidden layer, b^0 is the bias of the neuron at the output layer.

ARIMAX-NN is a hybrid method between ARIMAX as a linear method and NN as non-linear (Kusumaningrum *et al.*, 2023; Putera, 2020). The ARIMAX-NN hybrid model is represented by the following equation.

$$Z_t = Y_t + M_t$$

where Y_t is the result of the ARIMAX model and M_t is the result of the NN method forecasting.

2.5 Hybrid ARIMAX-SVR model

Support vector regression (SVR) is an algorithm that uses time series data patterns and kernel functions to predict nonlinear components of the data (Sihombing *et al.*, 2022). SVR is mapped into a high-dimensional feature space through a linear function hypothesis to exercise the structural risk minimization principle (Rubio & Alba, 2022). The kernel function has three parameters, namely cost (C), gamma (γ), and epsilon (ε) which are searched using the grid search method. This method consists of loose grid and finer grid. The SVR regression function equation is presented below.

$$N_t = \sum_{i-1}^{m} (\alpha_i - \alpha_i^*) \exp\left(-\gamma \left|\left|x_i - x\right|\right|^2\right) + b$$

with $0 < \alpha_i \leq C, 0 < \alpha_i^* \leq C$.

ARIMAX-SVR is a hybrid model combining ARIMAX, a linear method, and SVR, a nonlinear method, represented by following equation.

$$Z_t = Y_t + N_t$$

where Y_t is the result of the ARIMAX model and N_t is the result of the SVR method forecasting.

2.6 Hybrid reg-ARIMA model

The Reg-ARIMA model combines regression and ARIMA techniques to forecast data, accommodating various seasonal patterns with different periods. In this model, the error serves as the process variable for ARIMA, while a weight, acting as a regressor variable, enhances the regression component (Bell & Hillmer, 1983; Lin & Liu, 2002). The linear regression equation is written as.

$$O_t = x\beta + \varepsilon$$

Furthermore, the ARIMA model can be written simply as follows.

$$P_t = \phi_1 X_{t-1} + \phi_2 X_{t-1} + \cdots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} \\ -\theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Reg-ARIMA is a hybrid method between Regression and ARIMA. The Reg-ARIMA hybrid model is represented by the following equation.

$$Z_t = O_t + P_t$$

where O_t is the result of the Regression model and P_t is the result of forecasting the ARIMA method.

2.7 Forecasting accuracy

Forecasting accuracy is crucial in research as it gauges the quality and confidence of results, facilitating decision-making. Various methods, including MAE, MSE, RMSE, MAPE, SMAPE, and others, are employed to calculate forecasting accuracy (Hyndman & Koehler, 2006). In this research, RMSE and MAPE will be used to calculate forecasting accuracy.

RMSE (Root Mean Square Error) is the root mean square error. Here is the formula.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)^2}$$

MAPE (mean absolute percentage error) is a form of absolute average percentage error.

$$MAPE = 100 \times \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right|$$

With n being the number of forecast data, Z_t is the actual data and $\hat{Z}_t$ is the forecasted data.

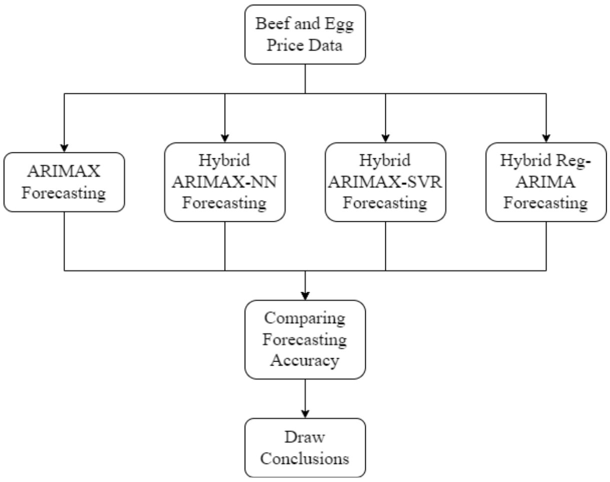

Figure 1. Forecasting design with four methods

3. Results and Discussion

The paper compares the performance of ARIMAX, ARIMAX-NN, ARIMAX-SVR, and Reg-ARIMA models through empirical assessment.

3.1 ARIMAX modeling

3.1.1 Hybrid ARIMAX modeling of beef price data in Central Java province

In the time series process, there are several important stages that must be met to obtain the ARIMAX model, one of which is data stationarity. The stationarity test conducted in this study uses the augmented Dickey Fuller (ADF) test. The results perform the data is not stationary. Therefore, a differencing process is carried out to stationarize the data. Then, identify the ARIMA model with ACF and PACF plots.

In the next process, namely the significance test of the model formed with dummy variables to the ARIMA model, it was found that the ARIMAX(1,1,0) model was the best model. The parameter estimation results show that the AR(1) parameters and dummy variables during Eid al-Fitr are significant parameters, so the best model formed is as follows.

$$\theta_p(B)(1-B)^d Y_t = \gamma_2 B_t$$
$$(1 + 0.847B)(1-B)^1 Y_t = 10.741.82B_t$$

Model was evaluated with RMSE and MAPE accuracy measures of 4246.497 and 0.0289, respectively.

3.1.2 Hybrid ARIMAX modeling of chicken egg price data in Central Java province

The stationarity test performed on this data uses the augmented Dickey Fuller (ADF) test. Results show that the p-value is smaller than the significant level of 0.05, meaning that the data is stationary. Then, the ARIMA model identification is carried out with ACF and PACF plots.

In the next process, namely the significance test of the model formed with dummy variables to the ARIMA model, it was found that the ARIMAX(1,0,0) model was the best model. The parameter estimation results show that the AR(1) parameters and dummy variables one month after Eid al-Fitr, one month before Christmas, and at Christmas are significant parameters, so the best model formed is as follows.

$$\theta_p(B)(1-B)^d Y_t = \gamma_3 B_{t+1} + \delta_1 C_{t-1} + \delta_2 C_t$$

$$(1-0.432B)(1-B)^0 Y_t = 1905.28B_{t+1} + 1481.364C_{t-1} + 4203.542C_t$$

$$(1-0.432B)Y_t = 1905.28B_{t+1} + 1481.364C_{t-1} + 4203.542C_t$$

Model was evaluated with RMSE and MAPE accuracy measures of 5682.446 and 0.2336, respectively.

3.2 Hybrid ARIMAX-NN modeling

The ARIMAX model's residuals showed nonlinear patterns, leading to a decision to switch to neural networks (NN). The NN model has input, hidden, and output layers. The input layer uses ARIMAX residuals, with neuron count based on significant lags. Experiments tested hidden layer neurons from 1 to 10, analyzing RMSE and MAPE values to find the best setup for each dataset.

3.2.1 Hybrid ARIMAX-NN modeling of beef price data in Central Java province

An ARIMAX model has been developed for beef price data in Central Java Province from May 2017 to August 2023. To address nonlinear residuals, the optimal Neural Network model is integrated into a hybrid ARIMAX-NN model. The input layer is determined based on the significant residuals' lag, set at 1. Experimentation with hidden layers reveals that the 9th hidden neuron yields the best results. The experimental stage indicates that a single epoch is sufficient for creating the most suitable model in this scenario. Then, the best neural network model selected is the NN(1,9,1) model with the architecture shown in Figure 2 (a).

The NN(1,9,1) residual ARIMAX model can be written as follows.

$$M_t = -0.11004 - 0.46191 f_1^{\boldsymbol{h}} + 0.08372 f_2^{\boldsymbol{h}} + 1.94204 f_3^{\boldsymbol{h}} + 0.78852 f_4^{\boldsymbol{h}} - 0.61873 f_5^{\boldsymbol{h}} - 1.1821 f_6^{\boldsymbol{h}} - 0.46489 f_7^{\boldsymbol{h}} - 0.45643 f_8^{\boldsymbol{h}} - 0.3525 f_9^{\boldsymbol{h}}$$

The hybrid ARIMAX-NN model can be written as follows.

$$Z_t = Y_t + M_t$$

$$Z_t = \frac{10.741{,}82B_t}{(1+0.847B)(1-B)^1} + (-0.11004 - 0.46191 f_1^{\boldsymbol{h}} + 0.08372 f_2^{\boldsymbol{h}} + 1.94204 f_3^{\boldsymbol{h}} + 0.78852 f_4^{\boldsymbol{h}} - 0.61873 f_5^{\boldsymbol{h}} - 1.1821 f_6^{\boldsymbol{h}} - 0.46489 f_7^{\boldsymbol{h}} - 0.45643 f_8^{\boldsymbol{h}} - 0.3525 f_9^{\boldsymbol{h}})$$

The ARIMAX-NN(1,9,1) model was evaluated with RMSE and MAPE accuracy measures of 4053.648 and 0.0278, respectively.

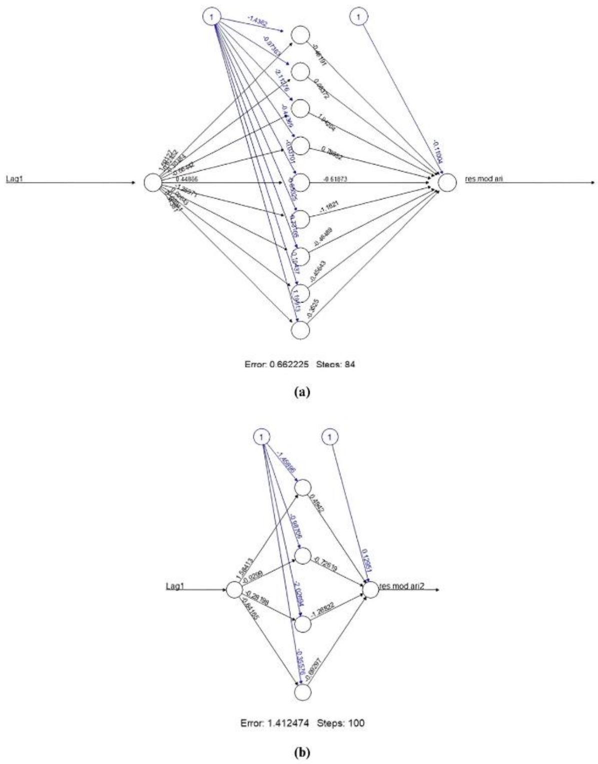

Figure 2. (a) ARIMAX-NN(1,9,1) architecture; (b) ARIMAX-NN(1,4,1) architecture

3.2.2 Hybrid ARIMAX-NN modeling of chicken egg price data in Central Java province

The ARIMAX model for Central Java Province's chicken egg price data from May 2017 to August 2023 has been created. To address nonlinear residuals, the optimal Neural Network model will be integrated to form a hybrid ARIMAX-NN model. The input layer is determined based on the number of lags with significant residuals, set at 1. Experimentation with hidden layers reveals that the 4th hidden neuron yields the best results. The experimental stage indicates that a single epoch is most suitable for this situation. Then, the best Neural Network model selected is the NN(1,4,1) model with the architecture shown in Figure 2 (b). The NN(1,4,1) residual ARIMAX model can be written as follows.

$$M_t = 0.12951 + 0.4942 f_1^h - 0.72619 f_2^h - 1.26822 f_3^h - 0.69297 f_4^h$$

The hybrid ARIMAX-NN model can be written as follows.

$$Z_t = Y_t + M_t$$

$$Z_t = \left(\frac{1905.28 B_{t+1} + 1481.364 C_{t-1} + 4203.542 C_t}{(1-0.432B)}\right) + (0.12951 + 0.4942 f_1^h - 0.72619 f_2^h - 1.26822 f_3^h - 0.69297 f_4^h)$$

The ARIMAX-NN(1,4,1) model was evaluated with RMSE and MAPE accuracy measures of 5617.843 and 0.2297, respectively.

3.3 Hybrid ARIMAX-SVR modeling

Hybrid ARIMAX-SVR modeling utilizes ARIMAX residuals with time lags. The optimal lag selection involves trial and error, revealing that the 14th lag minimizes errors and is consequently employed as input for the SVR model.

3.3.1 Hybrid ARIMAX-SVR modeling of beef price data in Central Java province

SVR parameters, namely cost (C), gamma (γ), and epsilon (ε) are determined by the grid search method. At the loose grid stage, the range of cost and gamma parameters are $2^{-9}, 2^{-7}, \ldots, 2^7, 2^9$, while the epsilon parameters are 0.01; 0.02;...;0.1. The best parameters for beef data at the loose grid stage are C=0.01, γ=4.76, ε=0.08. Furthermore, at the finer grid stage, the cost range is $2^{-6}, 2^{-6{,}25}, 2^{-6{,}5}, 2^{-6{,}75}, 2^{-5}$, the gamma parameter range is $2^2, 2^{2{,}25}, 2^{2{,}5}, 2^{2{,}75}, 2^3$. The parameter ε remains 0.01. So, the best parameters for the SVR model of beef price data are C=0.016,γ=4.76,ε=0.08. Then the SVR model with these parameters can be written as follows.

$$N_t = \sum_{t=1}^{61} (\alpha_t - \alpha_t^*) \exp\left(-4.76\left||x_t - x^2|\right|\right) + b,$$

with $0 < \alpha_t \leq 0.016, 0 < \alpha_t^* < 0.016$.

The ARIMAX-SVR hybrid model can be written as follows.

$$Z_t = Y_t + N_t$$

$$Z_t = \frac{10741.82 B_t}{(1+0.847B)(1-B)^1} + \sum_{t=1}^{61} (\alpha_t - \alpha_t^*) \exp\left(-4.76\left||x_t - x^2|\right|\right) + b$$

Next is to do the modeling with the combined ARIMAX-SVR method.

The ARIMAX-SVR model was evaluated with RMSE and MAPE accuracy measures of 4246.058 and 0.0289, respectively.

3.3.2 Hybrid ARIMAX-SVR modeling of chicken egg price data in Central Java province

SVR parameters, namely cost (C), gamma (γ), and epsilon (ε) are determined by the grid search method. At the loose grid stage, the range of cost and gamma parameters are $2^{-9}, 2^{-7}, \ldots, 2^7, 2^9$, while the epsilon parameters are 0.01; 0.02;...;0.1. The best parameters for chicken egg data at the loose grid stage are C = 4, γ = 64, ε = 0.06. Furthermore, at the finer grid stage, the cost range is $2^1, 2^{1{,}25}, 2^{1{,}5}, \ldots, 2^{2{,}75}, 2^3$, the gamma parameter range is $2^5, 2^{5{,}25}, 2^{5{,}5}, \ldots, 2^{6{,}75}, 2^7$. The parameter ε remains 0.06. So, the best parameters for the SVR model of beef price data are C=2.83, γ=64, ε=0.01. Then the SVR model with these parameters can be written as follows.

$$N_t = \sum_{t=1}^{61} (\alpha_t - \alpha_t^*) \exp\left(-64\left||x_t - x^2|\right|\right) + b,$$

with $0 < \alpha_t \leq 2.83, 0 < \alpha_t^* < 2.83$.

The hybrid ARIMAX-SVR model can be written as follows.

$$Z_t = Y_t + N_t$$

$$Z_t = \left(\frac{1905.28 B_{t+1} + 1481.364 C_{t-1} + 4203.542 C_t}{(1-0.432B)}\right) + \sum_{t=1}^{61} (\alpha_t - \alpha_t^*) \exp\left(-64\left||x_t - x^2|\right|\right) + b$$

Next is to do forecasting with the combined ARIMAX-SVR method.

The ARIMAX-SVR model was evaluated with RMSE and MAPE accuracy measures of 5577.318 and 0.22647, respectively.

3.4 Hybrid Reg-ARIMA modeling

3.4.1 Hybrid Reg-ARIMA modeling of beef price data in Central Java province

In forecasting using the Reg-ARIMA model, the dependent variable (Y_t) is beef price while the independent variable (X_t) is the weighting matrix. After the weighting matrix is used as an independent variable, then linear regression modeling is performed. The following model is formed.

$$O_t = 115231.5 - 700.7 A_{t-1} - 526.3 A_t + 348.5 A_{t+1} + 2826.9 B_{t-1} + 10968.5 B_t + 218.5 B_{t-1} - 191.5 C_{t-1} + 1098.7 C_t + 1224.3 C_{t+1}$$

The regression parameter significance test shows that the p-value is lower than the significant level of 0.05, meaning that the regression model formed has statistical significance. Next is the process of checking the stationarity of regression residual data. Based on the augmented Dickey Fuller (ADF) test results, the p-value is greater than the 0.05 significant level, meaning that the data is not stationary. The next process is to perform a differencing process to stationarize the data. Then, identify the ARIMA model with ACF and PACF plots. A temporary model is obtained in the Table 1. The table shows the results of assumption tests and significance tests with a significance level of 5% for each model.

In the next process, namely testing the significance of the model formed, it is obtained that the ARIMA (2,1,0) model is the best model, so the ARIMA model formed can be written as follows.

$$Y_t = -0.852X_{t-1} - 0.356X_{t-2} + e_t$$

The hybrid Reg-ARIMA model can be written as follows.

$$Z_t = Y_t + O_t$$

$$Z_t = (-0.852X_{t-1} - 0.356X_{t-2} + e_t) + (115231.5 \\ -700.7A_{t-1} - 526.3A_t + 348.5A_{t+1} \\ +2826.9B_{t-1} + 10968.5B_t + 218.5B_{t-1} \\ -191.5C_{t-1} + 1098.7C_t + 1224.3C_{t+1})$$

Next is to do forecasting with the Reg-ARIMA combined method.

The Reg-ARIMA model was evaluated with RMSE and MAPE accuracy measures of 4248.221 and 0.0319, respectively.

3.4.2 Hybrid Reg-ARIMA modeling of chicken egg price data in Central Java province

In forecasting using the Reg-ARIMA model, the dependent variable (Y_t) is the price of chicken eggs while the independent variable (X_t) is the weighting matrix. After the weighting matrix is used as an independent variable, linear regression modeling is then performed. The following model is formed.

$$O_t = 21805.1 - 621.6A_{t-1} + 31.7A_t + 854.9A_{t+1} \\ +1928.2B_{t-1} + 1003.2B_t + 2274.9B_{t-1} \\ +1004.9C_{t-1} + 4419,.2C_t + 1315.8C_{t+1}$$

The regression parameter significance test shows that the p-value is lower than the significant level of 0.05, meaning that the regression model formed has statistical significance. Next is the process of checking the stationarity of regression residual data. Based on the augmented Dickey Fuller (ADF) test results, the p-value is greater than the 0.05 significant level, meaning that the data is not stationary. The next process is to perform a differencing process to stationarize the data. Then, identify the ARIMA model with ACF and PACF plots. A temporary model is obtained in Table 2. The table shows the results of assumption tests and significance tests with a significance level of 5% for each model.

In the next process, namely testing the significance of the model formed, it is obtained that the ARIMA (0,1,2) model is the best model, so the ARIMA model formed can be written as follows.

$$Y_t = e_t - 0.0000013e_{t-1} - 0.048e_{(t-2)}$$

The hybrid Reg-ARIMA model can be written as follows.

$$Z_t = Y_t + O_t$$

$$Z_t = \left(e_t - 0.0000013e_{t-1} - 0.048e_{(t-2)}\right) \\ +(21805.1 - 621.6A_{t-1} + 31.7A_t \\ +854.9A_{t+1} + 1928.2B_{t-1} + 1003.2B_t \\ +2274.9B_{t-1} + 1004.9C_{t-1} \\ +4419.2C_t + 1315.8C_{t+1})$$

Table 1. ARIMA model estimation of beef regression residuals

Model	Residual white noise test	Residual normality test	AIC	*p* Value test result
ARIMA(1,1,1)	Fulfill	Do Not Fulfill	1119.765	Not Significant
ARIMA(1,1,0)	Fulfill	Fulfill	1122.812	Significant
ARIMA(0,1,1)	Fulfill	Do Not Fulfill	1120.048	Significant
ARIMA(2,1,2)	Fulfill	Do Not Fulfill	1123.764	Not Significant
ARIMA(2,1,0)	Fulfill	Fulfill	1120.865	Significant
ARIMA(0,1,2)	Fulfill	Do Not Fulfill	1119.928	Not Significant
ARIMA(2,1,1)	Fulfill	Do Not Fulfill	1121.761	Significant
ARIMA(1,1,2)	Fulfill	Do Not Fulfill	1121.762	Significant

Table 2. ARIMA model estimation of chicken egg regression residuals

Model	Residual white noise test	Residual normality test	AIC	*p* Value test result
ARIMA(1,1,2)	Fulfill	Fulfill	1055.315	Not Significant
ARIMA(2,1,1)	Fulfill	Fulfill	1054.116	Not Significant
ARIMA(2,1,0)	Fulfill	Fulfill	1065.073	Significant
ARIMA(2,1,2)	Fulfill	Fulfill	1055.708	Not Significant
ARIMA(0,1,2)	Fulfill	Fulfill	1053.372	Significant
ARIMA(1,1,1)	Fulfill	Fulfill	1054.254	Not Significant
ARIMA(0,1,1)	Fulfill	Fulfill	1054.631	Significant
ARIMA(1,1,0)	Fulfill	Fulfill	1069.774	Significant

Next is to do forecasting with the Reg-ARIMA combined method.

The Reg-ARIMA model was evaluated using RMSE and MAPE, yielding 5396.255 and 0.218, respectively.

3.5 Forecast results

After modeling and forecasting with the ARIMAX, ARIMAX-NN, ARIMAX-SVR, and Reg-ARIMA methods, the RMSE and MAPE values of the models formed will be compared to obtain the best method. The following is a comparison of the models formed.

The ARIMAX-NN method has the lowest RMSE and MAPE values for beef price data, making it the best model for forecasting beef prices in Central Java. Similarly, the Reg-ARIMA method outperforms others for chicken egg price data, making it the preferred choice for forecasting chicken egg prices in the same region. Forecast results for beef and chicken egg prices for the next 12 periods from September 2023 to August 2024 will be derived using the optimal method.

Table 3. Comparison of RMSE and MAPE of ARIMAX, ARIMAX-NN, ARIMAX-SVR, Reg-ARIMA models

Research methods	Data			
	Beef price		Chicken egg price	
	RMSE	MAPE	RMSE	MAPE
ARIMAX	4246.497	0.0289	5682.446	0.2336
ARIMAX-NN	4053.648	0.0278	5617.843	0.2297
ARIMAX-SVR	4246.058	0.0289	5577.318	0.2265
Reg-ARIMA	4248.221	0.0319	5396.255	0.218

4. Conclusions

The study found that ARIMAX-NN and Reg-ARIMA models are the top choices for predicting beef and chicken egg prices in Central Java. This information can assist breeders in devising more effective product and marketing strategies. The ARIMAX-NN results show an accuracy of 2.78%, while previous research by Kusumaningrum, Purnamasari, & Siringoringo (2023) ranged at 4%. For Reg-arima method, this study produced an accuracy of 21.8% while previous research by Nisa *et al.* (2020) was 24.96%. Using monthly time series data to determine weighting variables in dummy variables is deemed ineffective in forecasting food prices as it may alter the public demand pattern for these goods. Further research is advised to utilize closer time series, like daily data, for a more accurate forecasting of food prices.

Table 4. Beef price forecasting period September 2023 to August 2024

No	Period	Prediction
1	Sep-23	IDR 130,206
2	Oct-23	IDR 139,798
3	Nov-23	IDR 129,773
4	Dec-23	IDR 129,251
5	Jan-24	IDR 129,278
6	Feb-24	IDR 129,262
7	Mar-24	IDR 129,302
8	Apr-24	IDR 130,331
9	May-24	IDR 129,879
10	Jun-24	IDR 129,267
11	Jul-24	IDR 129,268
12	Aug-24	IDR 130,192

Table 5. Forecasting chicken egg prices for the period September 2023 to August 2024

No	Period	Prediction
1	Sep-23	IDR 28,474.35
2	Oct-23	IDR 28,862.59
3	Nov-23	IDR 28,971.62
4	Dec-23	IDR 32,342.11
5	Jan-24	IDR 28,622.11
6	Feb-24	IDR 27,959.61
7	Mar-24	IDR 29,965.16
8	Apr-24	IDR 29,008.52
9	May-24	IDR 30,915.66
10	Jun-24	IDR 28,862.59
11	Jul-24	IDR 28,862.59
12	Aug-24	IDR 28,862.59

Acknowledgements

Authors express gratitude to the reviewers and editors for their helpful feedback, which enhanced the manuscript. They also extend thanks to all external contributors who participated in the research and writing endeavors.

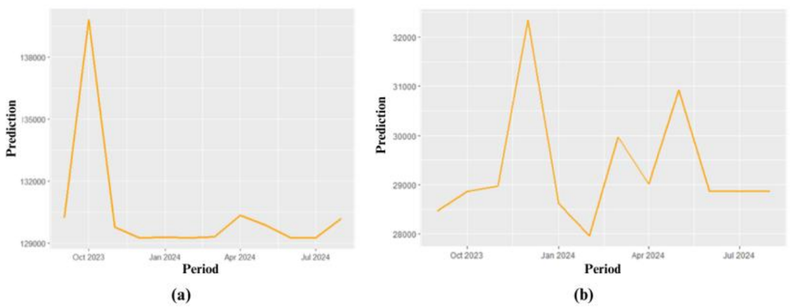

Figure 3. (a) Graph of beef price forecasting results; (b) Graph of chicken egg Price forecasting results

References

Anggraeni, W., Vinarti, R. A., & Kurniawati, Y. D. (2015). Performance comparisons between Arima and ARIMAX method in Moslem kids clothes demand forecasting: case study. *Procedia Computer Science*, *72*, 630–637.

Bell, W. R., & Hillmer, S. C. (1983). Modeling time series with calendar variation. *Journal of the American Statistical Association*, *78*(383), 526–534.

Chow, H. F. (2021). Short-term electricity grid maximum demand forecasting with the ARIMAX-SVR machine learning hybrid model. *The Hong Kong Institution of Engineers*, *28*(1), 22–30.

Gao, L., & Kling, G. (2005). Calendar effects in Chinese stock market. *Annals of Economics and Finance*, *6*(1), 75–88.

Hendikawati, P., Subanar, Abdurakhman, & Tarno. (2021). Non-stationary exchange rate prediction using soft computing techniques. *Songklanakarin Journal of Science and Technology*, *43*(2), 422–430.

Hillmer, S. C. (1982). Forecasting time series with trading day variation. *Journal of Forecasting*, *1*(4), 385–395.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at forecast-accuracy metrics for inter-mittent demand. *International Journal of Forecasting*, *22*, 679–688.

Intan, S. N., Zukhronah, E., & Wibowo, S. (2018). Peramalan banyaknya pengunjung Pantai Glagah menggunakan metode Autoregressive Integrated Moving Average Exogenous (ARIMAX) dengan efek variasi kalender [Forecasting the number of visitors to Glagah Beach using the autoregressive integrated moving average exogenous (ARIMAX) method with calendar variation effects]. *Indonesian Journal of Applied Statistics*, *1*(2), 70–78.

Jesica, H. P., Ispriyanti, D., & Tarno. (2019). Peramalan jumlah wisatawan yang berkunjung ke objek wisata di Jawa Tengah menggunakan variasi kalender Islam REGARIMA [Forecasting the number of tourists visiting tourist attractions in Central Java using Islamic calendar variation REGARIMA]. *Jurnal Gaussian*, *8*(3), 305–316.

Kusumaningrum, N., Purnamasari, I., & Siringoringo, M. (2023). Peramalan menggunakan model hybrid ARIMAX-NN untuk total transaksi pembayaran nontunai [Forecasting using hybrid ARIMAX-NN model for total cashless payment transactions]. *VARIANSI: Jurnal of Statistic and Its Application on Teaching and Research*, *5*(1), 1–14.

Lestari, D. W. (2023). *Peramalan harga bawang merah dan cabai merah menggunakan model ARIMAX (studi kasus: harga bawang merah dan cabai merah Indonesia periode Juli 2017 sampai dengan Januari 2023) [Forecasting shallot and red chilli prices using the ARIMAX model (case study: Indonesian shallot and red chilli prices for the period July 2017 to January 2023)]*. Yogyakarta, Indonesia: Islamic University of Indonesia.

Lin, J.-L., & Liu, T.-S. (2002). Modeling lunar calendar holiday effects in Taiwan. *Econometrics*, 1–29.

Liu, L. M. (1986). Identification of time series models in the presence of calendar variation. *International Journal of Forecasting*, *2*(3), 357–372.

Mills, T. C., & Coutts, J. A. (1995). Calendar effects in the London stock exchange FT-SE indices. *The European Journal of Finance*, *1*(1), 79–93.

Nisa, A. R., Tarno, & Rusgiyono, A. (2020). Peralaman harga cabai merah menggunakan model variasi kalender RegARIMA dengan moving holiday effect (studi kasus: harga cabai merah periode Januari 2012 sampai dengan Desember 2019 di Provinsi Jawa Barat) [Red chilli price forecasting using RegARIMA calendar variation model with moving holiday effect (case study: red chilli price from January 2012 to December 2019 in West Java Province)]. *Jurnal Gaussian*, *9*(2), 170–181.

Pavlenko, O., & Matvejevs, A. (2023). Comparison of linear and nonlinear models for forecasting of food commodity prices in Latvia. *Engineering for Rural Development*, 735–744.

Prastuti, M., Suhartono, & Salehah, N. (2018). Hybrid ARIMAX quantile regression method for dorecasting short term electricity consumption in East Java. *Journal of Physics: Conference Series*, *1008*(1), 1–10.

Putera, M. L. S. (2020). Peramalan transaksi pembayaran non-tunai menggunakan ARIMAX-ANN dengan konfigurasi kalender [Forecasting cashless payment transactions using ARIMAX-ANN with calendar configuration]. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, *14*(1), 135–146.

Qadrini, L., Asrirawan, Mahmudah, N., S, M. F., & Amri, I. F. (2021). Forecasting Bank Indonesia currency inflow and outflow using ARIMA, Time Series Regression (TSR), ARIMAX, and NN approaches in Lampung. *Jurnal Matematika. Statistika and Komputasi*, *17*(2), 166–177.

Rahmayani, L. (2013). *Model Arimax dan Sarimax untuk meramalkan data curah hujan [Arimax and Sarimax models for forecasting rainfall data]*. Bandung, Indonesia: State Islamic University Sunan Gunung Djati.

Riska, A. K., & Hasi, I. (2019). Peramalan nilai ekspor produk industri alas kaki menggunakan model ARIMAX dengan efek variasi kalender [Forecasting the export value of footwear industry products using the ARIMAX model with calendar variation effects]. *JSA: Jurnal Statistika Dan Aplikasinya*, *3*(2), 25–34.

Rubio, L., & Alba, K. (2022). Forecasting selected colombian shares using a hybrid ARIMA-SVR model. *Mathematics*, *10*, 1–21.

Sihombing, C. V. M., Martha, S., & Huda, N. M. (2022). Analisis metode hybrid ARIMA-SVR pada harga saham gabungan [Analysis of hybrid ARIMA-SVR method on joint stock price]. *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, *11*(3), 413–422.

Suhartono. (2006). Calendar variation model for forecasting time series with islamic calendar effect. *J urnal Matematika Sains dan Teknologi*, *7*(2), 85–94.

Suliyanto. (2008). *Teknik proyeksi bisnis: Teori dan aplikasi dengan Microsoft Excel [Business projection techniques: Theory and applications with microsoft excel]*, Yogyakarta, Indonesia: Andi.

Terui, N., & Dijk, H. K. van. (2002). Combined forecast from linear and nonlinear time series models. *International Journal of Forecasting*, *18*, 421–438.

Wei, W. W. S. (2006). *Time series analysis*. Hoboken, NJ: Pearson Addison Wesley.

Zhang, G. P. (2004). *Neural network in busines forecasting*. Hershy, PA: Idea Group Publishing.