International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 13; Issue: 5; May 2026
Website: www.ijrrjournal.com
Research Paper *E-ISSN: 2349-9788; P-ISSN: 2454-2237*

Modeling the Stock Price of PT Bank Negara Indonesia (Persero) Tbk Using Multivariable Local Polynomial Regression

Melani Ayu Azizah[1], Suparti[2], Diah Safitri[3]

[1,2,3]Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Corresponding Author: Diah Safitri

DOI: https://doi.org/10.52403/ijrr.20260571

ABSTRACT

Stocks are investment instruments with relatively high risk, and their prices form time series data. Forecasting stock price data helps investors make informed decisions. However, these data often exhibit complex, nonlinear patterns. Nonparametric regression offers a flexible approach that does not assume a specific functional form, allowing it to better adapt to data characteristics. This study models and forecasts the closing stock price of PT Bank Negara Indonesia (Persero) Tbk (BNI) using multivariable local polynomial regression. The data span from 1 December 2024 to 30 December 2025 and are split into 80% in-sample and 20% out-of-sample data. The in-sample data are used for stationarity testing, lag selection via the Partial Autocorrelation Function (PACF), model estimation using local polynomial regression with various bandwidths, and bandwidth selection based on the Generalized Cross Validation (GCV) criterion, followed by assessing model goodness-of-fit using the coefficient of determination. The out-of-sample data are used to evaluate the model's forecasting performance using the Mean Absolute Percentage Error (MAPE). The results show that a first-degree local polynomial regression model with the Epanechnikov kernel achieves the lowest Generalized Cross Validation (GCV) value of 270,710.862, a coefficient of determination of 82.33%, and an out-of-sample MAPE of 1.03%, indicating strong explanatory power and high forecasting accuracy for the BNI stock price.

Keywords: **S**tock price, Local polynomial regression, Generalized Cross Validation, Epanechnikov kernel, Forecasting.

INTRODUCTION

The development of the Indonesian capital market has shown a significant upward trend, as indicated by the increasing participation of the public in stock investment activities **[1].** Along with the growing interest in investment, stocks still carry risks such as capital loss due to price fluctuations. Therefore, stock price analysis is necessary to enable investors to make more optimal investment decisions. One of the prominent stocks in Indonesia is PT Bank Negara Indonesia (Persero) Tbk (BBNI), whose price movements are influenced by various internal company factors and market conditions.

The ARIMA model is widely used in time series forecasting. However, its parametric assumptions are often not suitable for capturing the complex characteristics of stock data. Therefore, a more flexible method is required. One approach that can be used is nonparametric regression, which does not require prior specification of a functional form [2].

Local polynomial regression is a nonparametric method that performs local approximation at each observation point by assigning weights based on a kernel function. The level of smoothness is controlled by a bandwidth parameter, which can be optimized using Generalized Cross Validation (GCV) [3]. This method is effective for data exhibiting clustered patterns [4]. In time series analysis, prediction problems can be formulated as regression problems using a flexible nonparametric approach without assuming a specific functional form, even in the presence of temporal dependence [5].

The identification of relevant lags in time series modeling is conducted using the Partial Autocorrelation Function (PACF), which measures the effect of each lag on the current value after controlling for other lags [6]. Several studies have shown that local polynomial regression provides good predictive performance for time series data, such as in forecasting global gold prices [7] and MDKA stock prices [8]. Based on preliminary analysis, BBNI stock price data exhibit significant PACF values up to lag 2 and show clustering patterns after data transformation. Therefore, this study aims to apply multivariable local polynomial regression with two lags to model the closing stock price of BBNI, considering the stationarity properties of the data.

MATERIALS & METHODS

Data Description

This study uses secondary data consisting of daily closing stock price of PT Bank Negara Indonesia (Persero) Tbk (BBNI), obtained from Yahoo Finance. The dataset includes weekday trading data from December 1, 2024 to December 30, 2025. The data are divided into 80% in-sample data for model development and 20% out-of-sample data for forecasting evaluation.

Stationary Testing

A time series is considered stationary when its mean and variance remain constant over time [15]. Stationarity in the mean is tested using the Dickey–Fuller test, as presented in Equation (1).

$$\Delta Z_t = \delta Z_{t-1} + \varepsilon_t \qquad (1)$$

where the null hypothesis $H_0: \delta = 0$ indicates non-stationarity. The null hypothesis is rejected when the p-value is less than α indicating stationarity in the mean. Otherwise, the series is considered non-stationary. If the series is non-stationary, differencing is applied as shown in Equation (2).

$$\Delta^d Z_t = (1-B)^d Z_t \qquad (2)$$

Stationarity in variance is assessed using the Box–Cox transformation [1], as presented in Equation (3).

$$Z_t^{*} = \begin{cases} \frac{Z_t^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln Z_t, & \lambda = 0 \end{cases} \qquad (3)$$

Lag Identification

Relevant lags are identified using the Partial Autocorrelation Function (PACF), as defined in Equation (4).

$$\phi_{p,p} = Corr(Z_t, Z_{t-p} \mid Z_{t-1}, \dots, Z_{t-(p-1)}) \qquad (4)$$

At a 5% significance level, PACF is considered statistically significant if it lies outside the confidence interval $\pm\frac{1.96}{\sqrt{n}}$. Significant lags are then used as predictor variables [9].

Model Specification

Local polynomial regression is a nonparametric method that approximates the regression function locally [3]. The general model is expressed in Equation (5).

$$Y_i = m(X_i) + \varepsilon_i \qquad (5)$$

For two predictor variables, the model becomes:

$$Y_i = m(X_{i1}, X_{i2}) + \varepsilon_i\,; \; i = 1,2,\dots,n \qquad (6)$$

This study considers polynomial degrees of 0 (local constant), 1 (local linear), and 2 (local quadratic) [10].

Parameter Estimation
Parameters are estimated using the Weighted Least Squares (WLS) method by minimizing the Weighted Residual Sum of Squares, as shown in Equation (7).

$$WRSS = \sum_{i=1}^{n}\{Y_i - m(\boldsymbol{X}_i)\}^2 \mathcal{K}_{\mathcal{H}}(\boldsymbol{X}_i - \boldsymbol{x}_0) \quad (7)$$

where $\mathcal{K}_{\mathcal{H}}(\boldsymbol{X}_i - \boldsymbol{x}_0)$ is a multivariable kernel function. The Epanechnikov kernel is used in this study. The parameter estimator is given by Equation (8).

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \quad (8)$$

Bandwidth Selection using GCV
Bandwidth plays a crucial role in controlling the smoothness of the estimated curve. The optimal bandwidth is determined using Generalized Cross Validation (GCV), as presented in Equation (9).

$$GCV(\mathcal{H}) = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - m(\boldsymbol{X}_i))^2}{\left(1 - \frac{1}{n}trace(\boldsymbol{H})\right)^2} \quad (9)$$

with

$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}$

The optimal bandwidth minimizes the GCV value, balancing bias and variance **[4]**.

Model Evaluation Metrics
Model performance is quantified using the coefficient of determination, as defined in Equation (10).

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(Y_t - \hat{Y})^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \quad (10)$$

which measures the proportion of variance explained by the model **[11]**.
Forecasting accuracy is evaluated using Mean Absolute Percentage Error (MAPE), as shown in Equation (11).

$$MAPE = \frac{\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|}{n} x\, 100\% \quad (11)$$

Lower MAPE values indicate higher forecasting accuracy **[12]**.

Modeling Procedure
The modeling process begins with testing the stationarity of the in-sample data. If necessary, transformations are applied to achieve stationarity. Significant lags are identified using PACF and used as predictor variables. The model is then estimated using multivariable local polynomial regression with the epanechnikov kernel, considering polynomial degrees of 0, 1, and 2 and various bandwidth values.
The optimal model is selected based on the minimum GCV value, and its goodness-of-fit is measured using the coefficient of determination. The selected model is then applied to out-of-sample data to evaluate forecasting performance using MAPE and to generate forecasts for future periods.
study.

RESULT
Observations/The data used in this study consist of daily closing price of BBNI stock from December 1, 2024 to December 30, 2025. The time series plot of the data is presented in Figure 1.

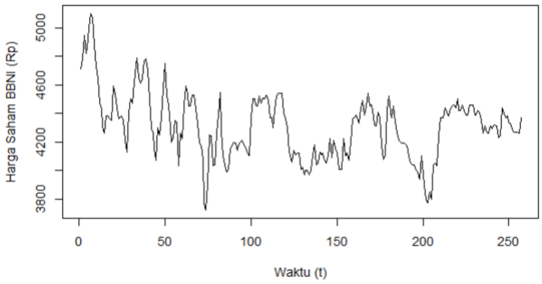

Figure 1: Time Series Plot of Data

The stationarity in variance was examined using the Box–Cox transformation. The estimated value of λ for the in-sample data was 1.19, indicating that the data were not yet stationary in variance. After applying the Box–Cox transformation, the value of λ became approximately 1, indicating that the transformed data satisfied the variance stationarity assumption.
The stationarity in mean was tested using the Dickey–Fuller test. The hypotheses used in this test are:
$H_0: \delta = 0$ (non-stationary)
$H_1: \delta < 0$ (stationary)
The test results are presented in Table 1.

Table 1. Dickey–Fuller Test Results

Data	τ	*p-value*	Decision
Transformed Data	-3,4575	0,01	Reject H_0

Based on the results, the p-value is less than the significance level $\alpha = 0.05$, indicating that H_0 is rejected. Therefore, the data is stationary in mean without requiring differencing. After achieving stationarity, the data were transformed into response and predictor variables based on significant lags identified from the PACF plot in Figure 2.

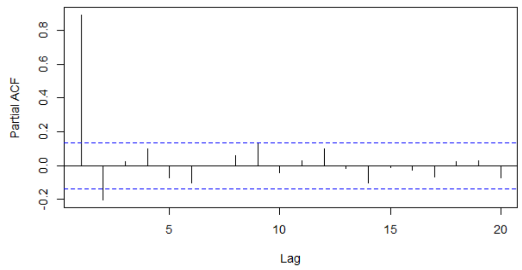

Figure 2: PACF Plot of BNI Transformation Stock Price

Based on Figure 2, two significant lags were identified. Therefore, response variable was defined as $Y^* = Z_t^*$ and predictor variables were defined as $X_1^* = Z_{t-1}^*$ and $X_2^* = Z_{t-2}^*$.

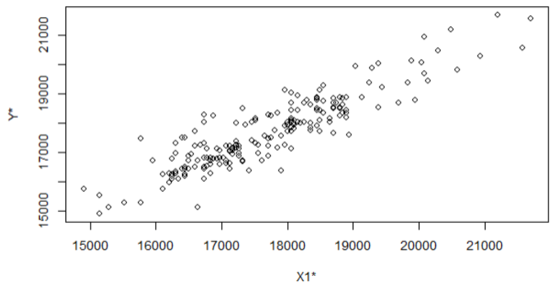

Figure 3: Scatterplot X_1* *and Y*

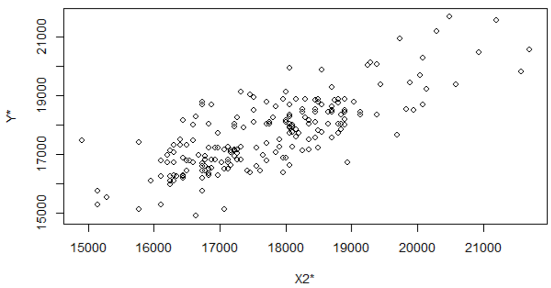

Figure 4: Scatterplot X_2* *and Y*

The relationships between the predictors and the response are illustrated using scatter plots in Figure 3 and Figure 4. The plots indicate that the data tend to cluster locally, suggesting that local polynomial regression is appropriate for capturing the relationship.
Model selection was conducted using a trial-and-error approach by considering polynomial degrees of 0, 1, and 2 with estimation based on the epanechnikov kernel and bandwidth selection using the Generalized Cross Validation (GCV) criterion. The local points x_{01}^* and x_{02}^* were selected within the range of predictor values, specifically between 15,000 and 22,000.
The bandwidth values were initially explored from 100 to 3,000 with increments of 100 and then refined with smaller increments to obtain the optimal value. The optimal model was selected based on the minimum GCV value. The results are presented in Table 2.

Table 2. Optimal Model Selection

Degree	x_{01}^*	x_{02}^*	Optimal h	GCV
0	17,300	18,500	1,800	1,508,162.264
1	17,500	17,600	1,560	270,710.862
2	16,300	18,400	2,604	3,609,205.355

The best model is the first-degree (local linear) model with the epanechnikov kernel, local points $x_{01}^* = 17{,}500$ and $x_{02}^* = 17{,}600$, and bandwidth 1,560, as it produces the minimum GCV value.
The estimated parameters of the selected model are presented in Table 3.

Table 3. Parameter Estimates

Beta	Estimates
β_0	17,486.453
β_1	1.098
β_2	-0.216

Based on these estimates, the local polynomial regression model in transformed form can be expressed as presented in Equation (12).

$$\hat{Y}_i^* = 17{,}486.453 + 1.098(X_{i1}^* - x_{01}^*) - 0.216\,(X_{i2}^* - x_{02}^*) \quad (12)$$

Substituting the predictor variables, the model can be rewritten as presented in Equation (13)

$$\hat{Z}_i^* = 17{,}486.453 + 1.098(Z_{t-1}^* - 17{,}500\,) - 0.216\,(Z_{t-2}^* - 17{,}600) \quad (13)$$

with

$$\hat{Z}_t = (\lambda \hat{Z}_t^* + 1)^{\frac{1}{\lambda}}\ ;\ Z_{t-1}^* = \frac{Z_{t-1}{}^{\lambda} - 1}{\lambda}\ ;\ Z_{t-2}^* = \frac{Z_{t-2}{}^{\lambda} - 1}{\lambda}\ ;\ \lambda = 1.19$$

Since the model is obtained in transformed form, the inverse Box–Cox transformation is applied to obtain predictions in the original scale.

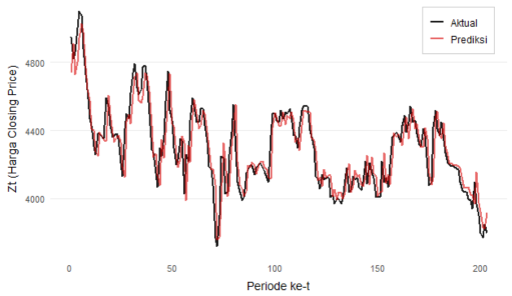

Figure 5: Plot of Actual Data and In-Sample Estimates of BBNI Stock Price

The comparison between actual and estimated in-sample data is shown in Figure 5. The coefficient of determination (R^2) is 0.82325, indicating that 82.325% of the variation in stock prices can be explained by the model. This value exceeds 0.67, suggesting that the model has strong explanatory power.

The comparison between actual and predicted out-sample data is shown in Figure 6.

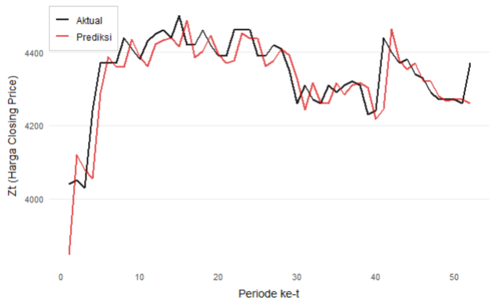

Figure 6: Plot of Actual Data and Out-of-Sample Estimates of BBNI Stock Price

The MAPE value is 1.027%, indicating a very high level of forecasting accuracy (MAPE < 10%).

Forecasting of BBNI stock prices is conducted for several future periods. The prediction can be performed flexibly, provided that data from the two previous periods are available. As an illustration, the forecasting of BBNI stock price is carried out starting from January 5, 2026, for the next five periods. The forecasting results for the next five periods are presented in Table 4.

Table 4. Forecasting Results

Nomor	Tanggal	Prediksi Harga Saham BBNI
1	5 Januari 2026	4.234
2	6 Januari 2026	4.233
3	7 Januari 2026	4.238
4	8 Januari 2026	4.243
5	9 Januari 2026	4.247

The model utilizes information from the previous two periods, enabling it to capture short-term dynamics and provide reliable forecasts for future stock price.

CONCLUSION

Based on the analysis results, the best multivariable local polynomial nonparametric regression model is obtained using the epanechnikov kernel with a polynomial degree of 1. The optimal model is constructed at the local points x01* = 17,500 and x02* = 17,600 with an optimal bandwidth of 1,560 determined based on the minimum Generalized Cross Validation (GCV) value of 270,710.862. The form of the best multivariable local polynomial nonparametric regression model can be seen in Equation (13).

The model is able to explain the variation in the data well, as indicated by a coefficient of determination (R^2) of 82.325%. The forecasting performance of the model is considered very good, based on a Mean Absolute Percentage Error (MAPE) of 1.027%.

Based on the best model obtained, stock price forecasting for BBNI was conducted for the next five days starting from January 5, 2026, using two lagged periods. The predicted values are 4,234; 4,233; 4,238; 4,243; and 4,247, respectively. The use of closing price information from the two previous periods enables the model to provide an overview of stock price movements in subsequent periods.

Declaration by Authors

Acknowledgement: None

Source of Funding: None

Conflict of Interest: No conflicts of interest declared.

REFERENCES

1. Bursa Efek Indonesia. 19 Juta Investor Pasar Modal dan 8 Juta Investor Saham Tercapai di Penutupan Bulan Inklusi Keuangan 2025 [Internet]. Jakarta: Bursa Efek Indonesia; 2025 [cited 2025 Nov 24]. Available from: https://www.idx.co.id/id/berita/siaran-pers/2488
2. Budiantara IN. Penelitian bidang regresi spline menuju terwujudnya penelitian statistika yang mandiri dan berkarakter. Surabaya: Institut Teknologi Sepuluh Nopember; 2011.
3. Fan J, Gijbels I. Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66. London: Chapman and Hall; 1996.
4. Suparti, Santoso R. Analisis data time series menggunakan model kernel: pemodelan data harga saham MDKA. Indonesian Journal of Applied Statistics. 2023; 6(1): 22-32. https://doi.org/10.13057/ijas.v6i1.79385
5. Härdle W. Applied Nonparametric Regression. Cambridge: Cambridge University Press; 1990.
6. Box GEP, Jenkins GM, Reinsel GC, *et al.* Time Series Analysis: Forecasting and Control. 5th ed. New York: John Wiley & Sons; 2015.
7. Hendrian J, Suparti, Prahutama A. Pemodelan harga emas dunia menggunakan metode nonparametrik polinomial lokal dilengkapi GUI R. Jurnal Gaussian. 2021; 10(4): 604-616. https://doi.org/10.14710/j.gauss.10.4.605-616
8. Fauzi FAN, Santoso R, Maruddani DAI. Pemodelan data time series menggunakan pendekatan regresi polinomial lokal pada data harga saham MDKA. Indonesian Journal of Applied Statistics. 2023; 6(2): 186-197. https://doi.org/10.13057/ijas.v6i2.80118
9. Wei WWS. Time Series Analysis: Univariate and Multivariate Methods. 2nd ed. Boston: Pearson/Addison Wesley; 2006.
10. Takezawa K. Introduction to Nonparametric Regression. New Jersey: John Wiley & Sons; 2006.
11. Gujarati DN, Porter DC. Basic Econometrics. 5th ed. New York: McGraw-Hill; 2009.
12. Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting: Methods and Applications. 3rd ed. New York: John Wiley & Sons; 1998.

How to cite this article: Melani Ayu Azizah, Suparti, Diah Safitri. Modeling the stock price of PT Bank Negara Indonesia (Persero) Tbk using multivariable local polynomial regression. *International Journal of Research and Review*. 2026; 13(5): 716-721. DOI: *https://doi.org/10.52403/ijrr.20260571*

******