Predicting Inflation in Indonesia Using Bi-predictors Semiparametric Model Based on Local Polynomial Estimator

Abdul Aziz[1,2*], Nur Chamidah[1] and Toha Saifudin[1]
[1]Doctoral Mathematics and Natural Sciences Study Program, Airlangga University, Surabaya, Indonesia
[2]Mathematics Study Program, Science and Technology Faculty, State Islamic University of Maulana Malik Ibrahim Malang, Indonesia

*abdulaziz@mat.uin-malang.ac.id

***Abstract*- Inflation is a condition of general and continuous increases in the prices of goods and services over a certain period of time. High and fluctuating inflation rates are a sign of economic instability. This fluctuating nature is due to factors that influence it, causing the relationship patterns in the data to not form a certain pattern and are also used for predictions. This research applies a semiparametric regression model, which combines parametric and nonparametric model, using a local polynomial estimator for inflation data with 2 factors that influence inflation, namely the Bank Indonesia (BI) interest rate in one previous month and the change rate of Money Supply (JUB) in one previous month. The local polynomial method estimates nonparametric functions by considering the local polynomial order and the optimum bandwidth value based on the lowest GCV value. In this research, a semiparametric regression model was obtained with an optimum bandwidth value of order 1, with high accuracy (MAPE 9.61%). The inflation predictions for September 2024, where the value is not yet known, with the BI interset rate and the change rate of JUB values in one previous month, using the model resulted in this research, the predicted value is 2.12%.**

Keywords—BI rate, inflation, JUB, local polynomial, semiparametric

I. INTRODUCTION

Many studies and research have been carried out on parametric regression for time series data. Inpartic ular, the ARIMA (Autoregressive Integrated Moving Average) model developed by [1], [2], [3], [4], [5], [6]. The ARIMA Box-Jenkins and ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) models were also carried out by [7]. The ARIMA model can be used for further data forecasting. Forecasting is the process of predicting future variables based on consideration of past data. Predictions Don't have to be able to provide exact answers certain about an event, but try to find an answer that is as close to the event as possible. The ARIMA model is applied for time series analysis, forecasting, and control. This technique is a projection technique that combines smoothing, regression and decomposition techniques. Because the ARIMA model is a combination of the AR and MA models, in this model the independent variables are the previous value (lag) of the dependent variable and the residual value of the previous period. Time series regression requires error assumptions to validate the model. A good error is the smallest possible error that is normally distributed and satisfies the properties of randomness. The ARIMA model is a parametric regression that takes into account error assumptions. Nonparametric regression is regression that does not use assumptions and rules like parametric regression. So, it can be said that nonparametric regression is more flexible than parametric regression.

Nonparametric regression has been widely used by several researchers with various estimator methods. Among them are [8], [9], [10], [11], [12], [13], [14] which use a single response nonparametric model with a local polynomial estimator. A more complex regression model, namely semiparametric, which combines parametric and non-parametric models, has been carried out by several researchers. Among them are adalah [23], [24], and [25] using local polynomial estimators in a single response semiparametric model. Local polynomials have several advantages, such as reducing asymptotic bias and producing good estimates [18]. Local polynomial estimators can be utilized by minimizing Weighted Least Squares (WLS). In local polynomial regression, the bandwidth determines how smooth the function is. The Generalized Cross Validation (GCV) method can be used to determine optimum bandwidth, which can be determined from the minimum GCV value. This can be seen from the work of [19].

2. MATERIAL AND METHODS

The bi-predictor semiparametric regression model, namely inflation as a response variable with two predictor variables, namely change rate of BI interest rate in the previous period as a parametric component with a linear function and change rate of JUB in the previous period as a nonparametric component, can be modeled with a semiparametric model as follows:

$$y_t = \beta_1 z_{t-1} + f\left(x_{t-1}\right) + \varepsilon_t \quad (1)$$

Bi-predictor semiparametric regression estimation cannot be carried out simultaneously, but one by one continuously. Starting from estimating parametric component parameters, then nonparametric functions based on local polynomial estimators which are solved using WLS with kernel function weighting. The function on the nonparametric component can be approximated via a Taylor series of order k, assuming the parameter values (β) on the parametric component are known:

$$\begin{aligned} f(x_{t-1}) &= \sum_{j=0}^{k} \frac{f^{(j)}(x_0)}{j!}(x_{t-1}-x_0)^j \\ &= \sum_{j=0}^{k} \lambda_j (x_{t-1}-x_0)^j \end{aligned} \quad (2)$$

So the equation (1) for all t can be written in the matrix equation form, namely,

$$Y = Z\beta + X\lambda + \varepsilon \quad (3)$$

The parameter λ in equation (3) depends on a point called a local point which is the data value for each observation sequentially, so that the first starting point will eliminate the first data, and so on. Assuming the estimator parameter β is known, parameter λ is estimated using Weighted Least Square (WLS) with a weighting kernel function K as diagonal matrix of Gaussian kernel function with bandwidth h,

$$K_h\left(x_{t-1}-x_0\right) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_{t-1}-x_0}{h}\right)^2\right) \quad (4)$$

by minimizing the weighted sum of squared error function,

$$\begin{aligned} S &= \left(Y^* - X\lambda\right)^T K \left(Y^* - X\lambda\right) \\ &= \left(Y^*\right)^T K Y^* - 2\lambda^T X^T K Y^* + \lambda^T X^T K X \lambda \end{aligned} \quad (5)$$

with

$$Y^* = Y - Z\beta = Y - Z\left(Z^T Z\right)^{-1} Z^T Y = Y - BY \quad (6)$$

and

$$B = Z\left(Z^T Z\right)^{-1} Z^T \quad (7)$$

which is the hat matrix in the parametric regression approach to Y.

The solution of equation (5) is,

$$\lambda = \left(X^T K X\right)^{-1} X^T K Y^* \quad (8)$$

So we get a function approach to nonparametric components,

$$f\left(X\right) = X\lambda = X\left(X^T K X\right)^{-1} X^T K Y^* = A^* Y^* \quad (9)$$

with,

$$A^* = X\left(X^T K X\right)^{-1} X^T K \quad (10)$$

which is the hat matrix in the nonparametric regression approach to Y*. Suppose Iy is a diagonal matrix of size (n-1) x (n-1), so by substituting equation (6) to equation (9) the final form of nonparametric component parameter estimation is obtained:

$$\begin{aligned} \lambda &= \left(X^T K X\right)^{-1} X^T K\left(Y - BY\right) \\ &= \left(X^T K X\right)^{-1} X^T K\left(I_Y - B\right)Y \end{aligned} \quad (11)$$

So we get a function approach to nonparametric components,

$$f\left(X\right) = X\lambda = AY \quad (12)$$

with,

$$A = X\left(X^T K X\right)^{-1} X^T K\left(I_Y - B\right) \quad (13)$$

which is the hat matrix in the nonparametric regression approach to Y.

From obtaining the form of estimation of the two parameters above, (7) and (15), a semiparametric model regression estimation form is obtained,

$$Y = Z\beta + X\lambda = BY + AY = CY \quad (14)$$

with,

$$C = A + B \quad (15)$$

which is the hat matrix in the semiparametric regression approach to Y.

In the local polynomial estimator, the initial value (x0) will be used as an estimation point which will create a local polynomial regression for surrounding data values with a bandwidth h. So the row value x0 in matrix C will become the hat matrix for a local polynomial around x0. This means that there will be as many local polynomials as the initial values used by matrix C which will become elements of each row in matrix C as a hat matrix for all local polynomials formed, namely M.

Apart from depending on the bandwidth value h, the parameter λ also depends on the polynomial order k. So that many estimated values of the parameter λ are obtained from repetition for a combination of the two (polynomial order k and bandwidth h). Therefore, to determine the optimum bandwidth value with a certain polynomial order, cross validation criteria are needed by selecting the minimum GCV value,

$$GCV(h)=\frac{1}{n-1}\sum_{t=2}^{n}\left[\frac{y_t^{*}-\sum_{j=0}^{k}\lambda_j\left(x_{t-1}-x_0\right)^j}{1-(n-1)^{-1}tr(M)}\right]^2 \qquad (16)$$

Furthermore, by obtaining the predicted Y values from equation (16) the GCV and MAPE values can be calculated for each repetition process of the combination above. So that an optimum combination (polynomial order and bandwidth value) can be obtained based on the smallest MAPE value. In this optimum combination, estimated parameter values are obtained which will form a semiparametric regression model.

III. RESULT

A. *Correlation Analysis*

Implementation of a semiparametric regression model based on a local polynomial estimator to Inflation data, BI interest rate in previous period (lag-1), and change rate of JUB in previous period (lag-1), require the help of statistical program applications. In this research, tools in the form of Open-Source Software R (OSS-R) were used. The following is a scatter plot image of the correlation matrix to determine the relationship pattern along with the correlation coefficient values between the three variables, namely:

Y : Observation data (factual) of Inflation
X1t-1 : Observation data (factual) of BI rate (lag-1)
dX2t-1 : Transformation data (differencing) of JUB (lag-1)

with

$$dX2_{t-1}=\frac{X2_{t-1}-X2_{t-2}}{X2_{t-2}} \qquad (17)$$

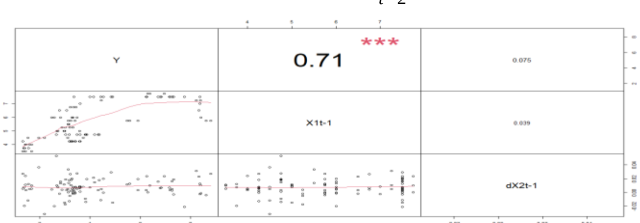

Figure 1. Scatter Plot of the correlation matrix for paired data on inflation, previous BI rate and previous change rate of JUB.

From Figure 1 it can be seen that the correlation coefficient value between the inflation variable in a period (Y) and the BI interest rate in the previous period (X1t-1) is 0.71 which is significant with an alpha of 0.1% (signed with thress stars) so that it can be said that there is a positive and quite strong relationship with an ascending linear relationship pattern. Meanwhile, the correlation coefficient value between the inflation variable in one period (Y) and change rate of JUB in the previous period (dX2t-1) is 0.075 which is not significant, so it can be said that there is no linear relationship. Meanwhile, the correlation coefficient value between BI interest rate in the previous period (X1t-1) and change rate of JUB in the previous period (dX2t-1) is 0.039 which is also not significant, so it can be said that there is no linear relationship between both variables. From the results of this correlation analysis, because two variables that is inflation and BI interest rate in previous period (Y and X1t-1) have a significant linear relationship and the two variables and two variables that is inflation and change rate of JUB in the previous period (Y and dX2t-1) do not have a linear relationship, they can be modeled semiparametrically, with Y is the response variable, X1t-1 is the predictor variable in the parametric component (here in after symbolized as Z), and dX2t-1 is the predictor variable in the nonparametric component (here in after symbolized as X). And there is no significant multicollinearity in both predictor variables (Z and X).

B. *Implementation of Local Polynomial*

Estimator Implementation of local polynomial estimator of polynomial order 1, 2, and 3 with many simulations bandwidth values on 100 training data (March 2013 to June 2021), produce bandwidth value 0.0001 with the smallest GCV 7.915212 on polynomial order 1. The parametric parameter estimator (β) that resulted is 0.75144 and every local point (x0) on data has nonpametric parameter estimator (λ) resulted. Take any period of data namely with X = -0.01453 as local point, nonpametric parameter estimator (λ) that resulted by local polynomial order 1 estimator is -0.75634 and 0.00003. So the semiparametric model with the local point of local polynomial order 1 estimator can be writen as:

$$\begin{aligned} y_t &= 0.75144\,z_{t-1}-0.75634+0.00003\left(x_{t-1}-(-0.01453)\right)\\ &= 0.75144\,z_{t-1}-0.75634+0.00003\left(x_{t-1}+0.01453\right) \qquad (18)\\ &= 0.75144\,z_{t-1}-0.75634+0.00003\,x_{t-1}+0.0000004359\\ &= -0.7563404359+0.75144\,z_{t-1}+0.00003\,x_{t-1}\end{aligned}$$

If resigned to the observation data and substitute equation (17), equation (18) become:

$$\begin{aligned} y_t &= -0.7563404359+0.75144\,x1_{t-1}+0.00003\,dx2_{t-1}\\ &= -0.7563404359+0.75144\,x1_{t-1}+0.00003\left(\frac{x2_{t-1}-x2_{t-2}}{x2_{t-2}}\right) \qquad (19)\end{aligned}$$

The results of nonparametric and semiparametric regression on local polynomial time-ordered data of order 1 with each data value as local point are shown in Figure 2 and 3.

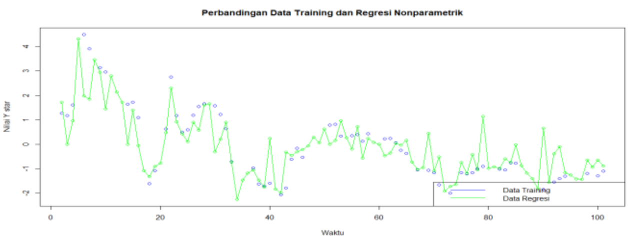

Figure 2. Comparison of time-ordered data with nonparametric regression.

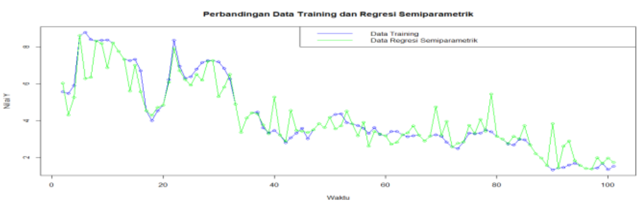

Figure 3. Comparison of time-ordered data with semiparametric regression.

C. *Prediction of inflation*

The optimum bandwidth value obtained from estimating semiparametric regression parameters can be used to estimate semiparametric model to the testing data (July 2021 to August 2024). With the bandwidth optimum value 0.0001 on polynomial order 1 on the testing data, the MAPE value resulted 9.61%, so it can be said that the model is high accuracy because less than 10%. The parametric parameter estimator (β) that resulted is 0.64853 and every local point (x_0) on data has nonpametric parameter estimator (λ) resulted. Take the last period of data (August 2024) with X = -0.00614 as local point, nonpametric parameter estimator (λ) that resulted by local polynomial order 1 estimator is -1.93242 and 0. So the semiparametric model with the local point of local polynomial order 1 estimator can be writen as:

$$y_t = 0.64853 z_{t-1} - 1.93242 + 0\left(x_{t-1} - \left(-0.00614\right)\right) \\ = 0.64853 z_{t-1} - 1.93242 \quad (20)$$

The results of nonparametric and semiparametric regression on local polynomial time-ordered data of order 1 with each data as local point are shown in Figure 4 and 5.

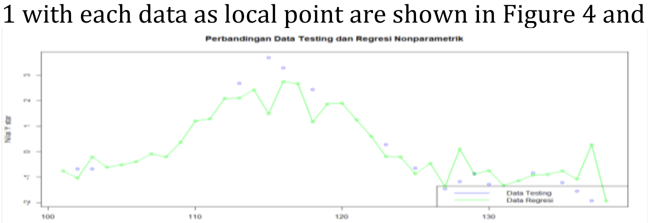

Figure 4. Comparison of time-ordered data with nonparametric regression.

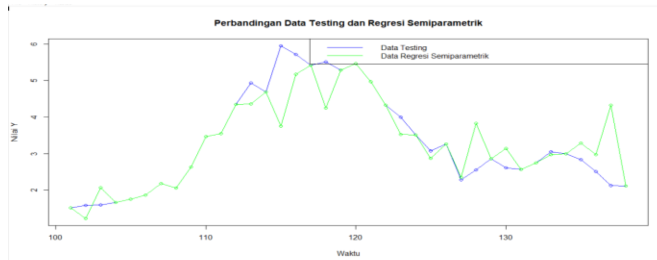

Figure 5. Comparison of time-ordered data with semiparametric regression.

The prediction for inflation value on September 2024, where the inflation value data is not yet known, the BI rate for August 2024 is 6.25%. Using the semiparametric regression model in equation (20) the inflation prediction value is obtained in July 2024:

$$y_t = 0.64853(6.25) - 1.93242 = 2.12 \quad (21)$$

IV. CONCLUSION

Implementation of the bi-predictor semiparametric model using a local polynomial estimator approach to the inflation rate in Indonesia with two predictors, BI interest rate in previous period and change rate of JUB in previous period month, resulted in a semiparametric regression model with an optimum bandwidth value is 0.00001 of polynomial order 1 with high accuracy. Predictions for September 2024, where the inflation value data is not yet known, using semparametric based on local polynomial order 1 estimator, the obtained value is 2.12%.

REFERENCES

[1] F. Z. Habbab and M. Kampouridis, "An in-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs," *Expert Syst Appl*, vol. 235, 2024, doi: 10.1016/j.eswa.2023.121102.

[2] V. P. Ariyanti and Tristyanti Yusnitasari, "Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, 2023, doi: 10.29207/resti.v7i2.4895.

[3] M. Abdelaziz, A. Ahmed, A. Riad, G. Abderrezak, and A. A. Djida, "Forecasting daily confirmed COVID-19 cases in Algeria using ARIMA models," *Eastern Mediterranean Health Journal*, vol. 29, no. 7, 2023, doi: 10.26719/emhj.23.054.

[4] C. D. Setiawan, W. Sulandari, and Y. Susanti, "Peramalan Harga Saham PT Unilever Indonesia Menggunakan Metode Hibrida ARIMA-Neural

Network," *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, vol. 7, no. 1, 2023, doi: 10.30998/semnasristek.v7i1.6270.

[5] S. R. Gusman, S. Suparti, and A. Rusgiyono, "Perbandingan Model Arima Dengan Model Nonparametrik Polinomial Lokal Dan Spline Truncated Untuk Peramalan Harga Minyak Mentah West Texas Intermediate (WTI) Dilengkapi GUI R," *Jurnal Gaussian*, vol. 12, no. 1, 2023, doi: 10.14710/j.gauss.12.1.20-29.

[6] R. Pratiwi and H. Yundari, "Model Arima Semiparametrik," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 11, no. 1, pp. 129–138, 2022.

[7] J. Iqbalullah and W. S. Winahju, "Peramalan Jumlah Penumpang Pesawat Terbang di Pintu Kedatangan Bandar Udara Internasional Lombok dengan Metode ARIMA Box-Jenkins, ARIMAX, dan Regresi Time Series," *Jurnal Sains dan Seni ITS; Vol 3, No 2 (2014); D212-D21*, Jul. 2014, [Online]. Available: http://ejurnal.its.ac.id/index.php/sains_seni/article/view/8138

[8] F. Bravo, "Local polynomial estimation of nonparametric general estimating equations," *Stat Probab Lett*, vol. 197, 2023, doi: 10.1016/j.spl.2023.109805.

[9] M. Ulya, N. Chamidah, and T. Saifudin, "Prediction Of pH And Total Soluble Solids Content Of Mango Using Biresponse Multipredictor Local Polynomial Nonparametric Regression," *Communications in Mathematical Biology and Neuroscience*, vol. 2023, 2023, doi: 10.28919/cmbn/7941.

[10] S. E. Ahmed, D. Aydın, and E. Yılmaz, "Penalty and Shrinkage Strategies Based on Local Polynomials for Right-Censored Partially Linear Regression," *Entropy*, vol. 24, no. 12, 2022, doi: 10.3390/e24121833.

[11] Y. Wang, Y. Wu, and S. S. Du, "Near-Linear Time Local Polynomial Nonparametric Estimation with Box Kernels," *INFORMS J Comput*, vol. 33, no. 4, 2021, doi: 10.1287/ijoc.2020.1021.

[12] V. Fibriyani and N. Chamidah, "Prediction of Inflation Rate in Indonesia Using Local Polynomial Estimator for Time Series Data," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1776/1/012065.

[13] N. S. Rahmi, "Peramalan Inflow Uang Kartal Bank Indonesia KPW Tasikmalaya Jawa Barat Dengan Metode Klasik Dan Modern," *Jurnal Statistika Universitas Muhammadiyah Semarang; Vol 8, No 2 (2020): Jurnal Statistika; 166-174 ; 2528-1070 ; 2338-3216*, Jul. 2020, [Online]. Available: https://jurnal.unimus.ac.id/index.php/statistik/article/view/6412

[14] N. Chamidah and B. Lestari, "Estimation Of Covariance Matrix Using Multi-Response Local Polynomial Estimator For Designing Children Growth Charts: A Theoretically Discussion," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. 2019. doi: 10.1088/1742-6596/1397/1/012072.

[15] V. Fibriyani, N. Chamidah, and T. Saifudin, "Modeling Case Fatality Rate Of COVID-19 In Indonesia Using Time Series Semiparametric Regression Based On Local Polynomial Estimator," *Commun. Math. Biol. Neurosci.*, vol. 2024, p. Article–ID, 2024.

[16] S. E. Ahmed, D. Aydin, and E. Yilmaz, "Semiparametric Time-Series Model Using Local Polynomial: An Application on the Effects of Financial Risk Factors on Crop Yield," *Journal of Risk and Financial Management*, vol. 15, no. 141, pp. 1–12, Mar. 2022, doi: 10.3390/jrfm15030141.

[17] G. Aryal, M. F. Gabrielli, and Q. Vuong, "Semiparametric Estimation of First-Price Auction Models," *Journal of Business & Economic Statistics*, vol. 39, no. 2, pp. 373–385, Apr. 2021, doi: 10.1080/07350015.2019.1665530.

[18] A. H. Welsh and T. W. Yee, "Local regression for vector responses," *J Stat Plan Inference*, vol. 136, no. 9, pp. 3007–3031, Sep. 2006, doi: 10.1016/j.jspi.2004.01.024.

[19] N. G. Zia, S. Suparti, and D. Safitri, "Pemodelan Regresi Spline Menggunakan Metode Penalized Spline Pada Data Longitudinal," *Jurnal Gaussian*, vol. 6, no. 2, 2017, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/article/view/16951