

Quantile Regression Neural Network Model For Forecasting Consumer Price Index In Indonesia

Dwi Rantini, Made Ayu Dwi Octavanny, Rumaisa Kruba, Heri Kuswanto*, Kartika Fithriasari

Department of Statistics, Faculty of Mathematics, Computing, and Data Science

Institut Teknologi Sepuluh Nopember Surabaya

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: heri_k@statistika.its.ac.id

Abstract— The main purpose of time series analysis is to obtain the forecasting result from an observation for future values. Quantile Regression Neural Network is a statistical method that can model data with non-homogeneous variance with artificial neural network approach that can capture nonlinear patterns in the data. Real data that allegedly have such characteristics is Consumer Price Index (CPI). CPI forecasting is important to assess price changes associated with cost of living as well as identifying periods of inflation or deflation. The purpose of this research is to compare several method of forecasting CPI in Indonesia. The data used in this study during January 2007 until April 2018 period. QRNN method will be compared with Neural Network with RMSE evaluation criteria. The result is QRNN is the best method for forecasting CPI with RMSE 0.95.

Keywords—Consumer Price Index, Neural Network, Nonlinearity, Quantile Regression

I. INTRODUCTION

One of the statistical methods that is used for data analysis is time series analysis. The main purpose of time series analysis is to obtain the forecasting result from an observation for future values. Common modern forecasting method is Neural Network (NN). The advantages of NN method are can be used for non-linear data approach and universal approach with high accuracy and without assumptions [1].

It is common in financial sector, that the data for forecasting have non-homogenous variance. One of the methods for forecast with non-homogenous variance is Quantile Regression (QR). QR method does not need residual assumptions like Ordinary Least Square (OLS) method, such as identic, independence, normally distributed, and there is no heteroscedasticity [2]. Forecasting using QR will produce interval forecasting result.

In this research, QR will be applied in Customer Price Index (CPI) in Indonesia because it assumed to have heteroscedasticity problem. CPI is a measure that examines the weighted average of prices of a basket of consumer goods and services, such as transportation, food and medical care. CPI forecasting is important to assess price changes associated with cost of living as well as identifying periods of inflation or deflation [3]. Previous study about forecasting CPI with ARIMA model has been done in Bandar Lampung [4] and Rwanda [5].

Accuracy of CPI is the most important thing in order to make informed decisions about the economy. In according to increase the accuracy, this research will used hybrid method. In general, the combination of two methods will be better

than using one method [6]. QR method will be combined with Neural Network or known as Hybrid Quantile Regression Neural Network. This method has been used to estimating conditional density of multi-period return [7]. The result is Quantile Regression Neural Network better than GARCH when using empirical distribution and tend to have same performance with GARCH when using Gaussian distribution. Forecasting using combination of Quantile Regression and Neural Network or later known as Robust Neural Network (RNN) in the credit portfolio data gives results that the method is more resistant to outlier when compared with linear regression or spline regression [8].

II. LITERATURE STUDY

A. Terasvirta Nonlinearity Test

Terasvirta nonlinearity test is conducted to find out whether the data are linear or nonlinear patterns. The Terasvirta test uses the F test and available tribes result from the Taylor series expansion [9]. The hypothesis for the Terasvirta nonlinearity test is as follows.

H_0 : $f(x)$ is a linear function in x or linear model

H_1 : $f(x)$ is a nonlinear function in x or non-linear model

The procedure for obtaining F test statistic values is as follows.

1. Regress Y_t with $Y_t, Y_{t-1}, Y_{t-2}, Y_t, \dots, Y_{t-k}$, and calculate the residual \hat{u}_t as well as calculate the sum of residual squares

$$SSR_0 = \sum_{t=1}^r \hat{u}_t^2 \quad (1)$$

2. Regress $\hat{\alpha}_t$ with $Y_t, Y_{t-1}, Y_{t-2}, Y_t, \dots, Y_{t-k}$, and m additional predictors obtained from the Taylor series expansion, then calculating the residual \hat{v}_t as well as calculate the sum of residual squares.

$$SSR = \sum_{t=1}^r \hat{v}_t^2 \quad (2)$$

3. Calculate the statistic F test with the following formula:

$$F = \frac{(SSR_0 - SSR) / m}{SSR / (n - p - 1 - m)} \quad (3)$$

with n is the number of observations and p is a number of order. H_0 refused if $F > F_{(m, n-p-1-m)}$ which means that the model is not linear.

B. Quantile Regression

Quantile Regression (QR) was first introduced by Koenker and Bassett (1978). This approach assumes the various quantile functions of a Y distribution as a function of X . Quantile Regression is useful if the distribution of data is not homogeneous (heterogeneous) and not standardized as is not symmetric, there is a tail in distribution, or truncated distribution. Here is the pinball loss function of quantile regression [10].

$$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases} \quad (4)$$

With $0 < \tau < 1$. For example, the predictors $x_i(t)$, $i = 1, \dots, I$, slope coefficient m_i , and intercept b , linear regression equation for \hat{y}_t conditional on the τ quantile are as follows

$$Y_t(t) = \sum_{i=1}^I m_i x_i(t) + b \quad (5)$$

The coefficient of the equation can be estimated by minimizing the quantile regression error function as follows

$$E_\tau = \frac{1}{n} \sum_{t=1}^N \rho_\tau(Y(t) - \hat{Y}_\tau(t)), \quad (6)$$

$Y(t)$ being the response observed at time $t = 1, \dots, n$.

C. Neural Network

Neural Network is a technique in machine learning that has been developed as a generalization of the mathematical model of the biological nervous system. Learning methods in neural networks can be classified into three, namely supervised learning, unsupervised learning, and reinforcement learning. The neural network model commonly used in forecasting is feed forward neural network [1]. The following is example architecture model of feed forward neural network with three layer.

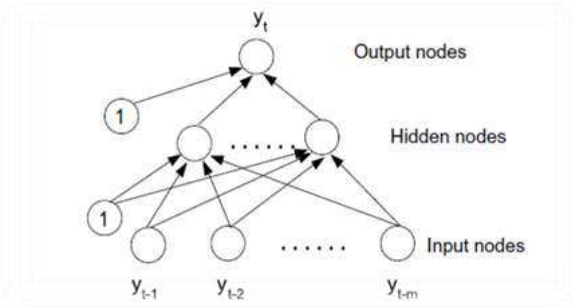


Fig. 1. Feed Forward Neural Network with Three Layers

The relationship between output Y_t and input $Y_{t-1}, Y_{t-2}, Y_t, \dots, Y_{t-k}$ is depicted in the following mathematical equation:

$$Y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g(\beta_{0j} + \sum_{i=1}^p \beta_{ij} Y_{t-i}) + \alpha_t \quad (7)$$

α_j ($j = 0, 1, 2, \dots, q$) and β_{ij} ($i = 0, 1, 2, \dots, p; j = 1, 2, \dots, q$) is a model parameter or often referred to as a weight, p is the number of input nodes, and q is the number of hidden nodes. The usual function used in the hidden layer is the logical function or tangent hyperbolic. Equation (7) establishes a nonlinear mapping of the observed past value

$(Y_{t-1}, Y_{t-2}, Y_t, \dots, Y_{t-k})$ to the future value of (Y_t) with the following equation,

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_t, \dots, Y_{t-k}, w) + \alpha_t \quad (8)$$

where w is the vector of all parameters and f is the function determined from the network structure and the weights. Thus, the neural network as in the picture is equivalent to the nonlinear autoregressive model [13].

D. Quantile Regression Neural Network

Quantile Regression Neural Network is a combination of Quantile Regression and Neural Network methods. For example known predictor $x_i(t)$ and response $y(t)$. The output of the hidden-layer node j is as follows [10].

$$g_j(t) = \tanh\left(\sum_{i=1}^I x_i(t) w_{ij}^{(h)} + b_j^{(h)}\right) \quad (9)$$

$w_{ij}^{(h)}$ is the weight of the hidden-layer while $b_j^{(h)}$ is the bias of the hidden-layer. The estimation of $\hat{y}_\tau(t)$ conditional quantile for $k - \tau$ is as follows

$$\hat{y}_\tau(t) = f\left(\sum_{j=1}^J g_j(t) w_j^{(o)} + b^{(o)}\right), \quad (10)$$

$w_j^{(o)}$ is the weight of the output-layer and $b^{(o)}$ is biased at the output-layer whereas $f(\cdot)$ is a function of the output-layer. Loss function is given in the following equation

$$\rho_\tau^{(a)}(u) = \begin{cases} \tau h(u) & \text{if } u \geq 0 \\ (\tau - 1)h(u) & \text{if } u < 0 \end{cases} \quad (11)$$

where u is the residual of the parameter estimate and $h(u)$ is Huber norm with the following equation

$$h(u) = \begin{cases} \frac{u^2}{2\varepsilon} & \text{if } 0 \leq |u| \leq \varepsilon \\ |u| - \frac{\varepsilon}{2} & \text{if } |u| \geq \varepsilon \end{cases} \quad (12)$$

With ε is the threshold whose value is determined. The error function to be optimized is given in the following equation

$$E_\tau^{(\alpha)} = \sum_{t=1}^n \frac{1}{n} \sum_{i=1}^N \rho_\tau(Y(t) - \hat{Y}_\tau(t)), \quad (13)$$

E. Consumer Price Index (CPI)

The Consumer Price Index (CPI) is a measure to examines the weighted average of prices of a basket of consumer good and services, such as transportation, food and medical care. It is calculated by taking price changes for each item in the predetermined basket of goods and average value of them. Changes in the CPI are used to assess price changes associated with the living cost. The CPI can also be used as a deflator for other economic factors, including retail sales, hourly/weekly earnings and the value of a consumer's dollar to find its purchasing power [3].

The BLS records about 80,000 items each month by calling or visiting retail stores, service establishments (such as cable providers, airlines, car and truck rental agencies), rental units and doctors' offices across the country in order to get the best outlook for the CPI. The formula used to calculate the Consumer Price Index for a single item.

$$CPI = \frac{\text{Cost of Market Basket in Given Year}}{\text{Cost of Market Basket in Base Year}} \times 100 \quad (14)$$

The base year is determined by the Bureau of Labor Statistics (BLS). CPI data for the years 2016 and 2017 were based on surveys collected in 2013 and 2014 [11].

III. METHODOLOGY

This research uses Consumer Price Index (CPI) data, which is obtained from Indonesia Central Bureau of Statistics. The period of data is January 2007 until April 2018, with training data in January 2007 until December 2017 and testing data in January 2018 until April 2018.

The step of this research as follows:

1. Describe the characteristic of CPI data.
2. Modeling using Neural Network with significant lag in PACF from stationary data as input variable and CPI as output variable.
3. Modeling using Quantile Regression Neural Network with two kinds of model.
 - a. One-step model using significant lag in PACF from stationary data as input variable and CPI as output variable.
 - b. Using significant lag in PACF from residual of the best Neural Network model as input variable and residual as output variable (hybrid method).
4. Comparing the performance of Neural Network and Quantile Regression Neural Network using RMSE evaluation criteria.
5. Forecasting CPI in May 2018 until December 2018.
6. Draw a conclusion.

IV. ANALYSIS AND DISCUSSION

A. Data Description

Before analyzing the data, here is the time series plot for CPI data.

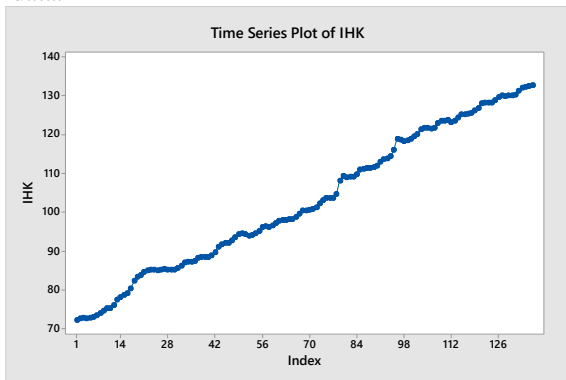


Fig. 2. Time Series Plot of CPI

From the Fig. 2, it can be seen that the CPI data is not stationary in the mean. In order to stationary of CPI data in mean then must be done differencing with first order. This stationary data to be analyzed using NN and QRNN. To be able to use the NN and QRNN methods, the data should not normally distributed and not linear. For CPI data, here is the normality test using Kolmogorov Smirnov and linearity test using Terasvista test.

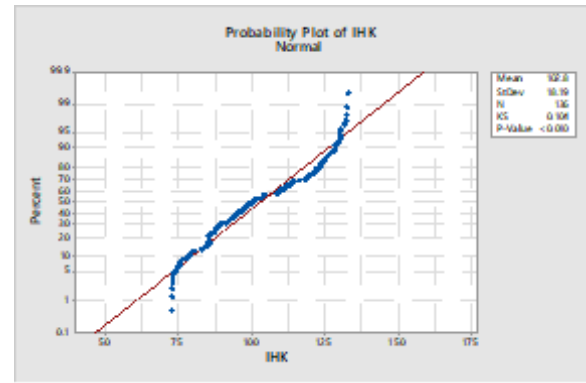


Fig. 3. Probability Plot of CPI

Based on Fig. 3, by using the level of significance $\alpha = 5\%$, then reject H_0 , meaning that the CPI data is not normally distributed. For linearity test, using the level of significance $\alpha = 5\%$, the result of χ^2 is 35.27, then reject H_0 , meaning that the CPI data is not linear.

B. Modeling Using Neural Network

In the NN method, the input variable is a significant lag of PACF from the already stationary input sequence, while the output variable is the CPI data.

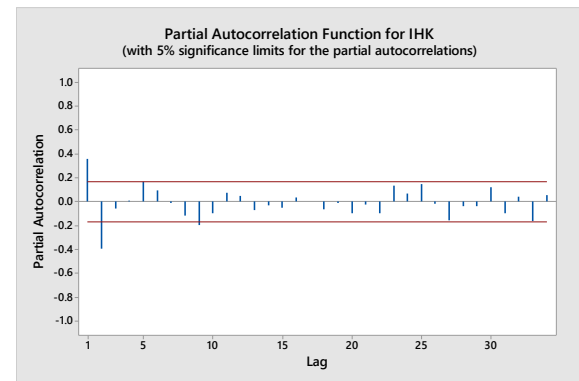


Fig. 4. PACF of Stationary CPI

From the PACF in Fig. 4 above, it can be seen that the significant lag is lag 1, 2 and 5, then this lag will be the input variable. With input variables lag 1, lag 2 and lag 5, CPI data as output variables, normalized transformation, and sigmoid activation function, the following results are obtained.

TABLE I. RMSE NEURAL NETWORK METHOD

Number of Node	Root Mean Squared Error (RMSE)	
	Training	Testing
1	6.98	2.27
2	6.99	1.62
3	5.32	1.07
4	6.62	1.77
5	7.16	2.18
6	6.76	1.71
7	7.20	1.84
8	6.36	2.33
9	6.18	1.22
10	9.22	2.98

Based on the RMSE value from Table I, the smallest RMSE value is NN with 3 nodes, and the architecture layer as follows in Fig 5.

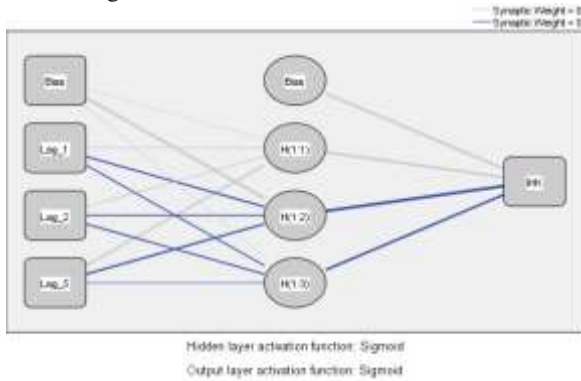


Fig. 5. PACF of Stationary CPI

Parameter estimation using NN as in Table II.

TABLE II. PARAMETER ESTIMATION USING NEURAL NETWORK

Predictor		Predicted			
		Hidden Layer 1			Output Layer
		H(1:1)	H(1:2)	H(1:3)	
Input Layer	(Bias)	0.229	4.270	.381	
	Lag_1	0.994	-1.360	-1.148	
	Lag_2	1.059	-1.078	-1.312	
	Lag_5	1.584	-1.529	-.509	
Hidden Layer 1	(Bias)				3.570
	H(1:1)				3.482
	H(1:2)				-6.761
	H(1:3)				-1.991

C. Modeling Using Quantile Regression Neural Network

Modeling using Quantile Regression Neural Network with two kinds of model as follows.

1. One-step model using significant lag in PACF from stationary data as input variable and CPI as output variable.

In the QRNN method, same as NN method, the input variable is a significant lag of PACF from the already stationary input sequence, while the output variable is the CPI data. Same as the NN method, the CPI data must be transformed normalized. RMSE for some quantiles are obtained as in Table III.

TABLE III. RMSE QRNN METHOD (ONE STEP)

Quantile	RMSE
0.025	0.83
0.5	0.95
0.075	1.31

2. Using significant lag in PACF from residual of the best Neural Network model as input variable and residual as output variable (hybrid method).

For hybrid method NN-QRNN required input and output variables. the input variable is given by a significant lag of residual PACF NN3, while the output variable is residual NN3.

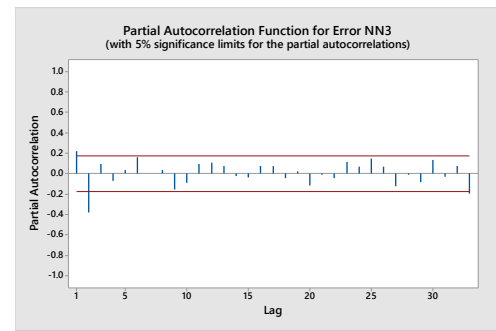


Fig. 6. PACF of Stationary CPI

Based on PACF residual NN3 in Fig. 6, results a significant lag of lag 1 and lag 2 as input variable. Same with NN method, the CPI data must be transformed normalized. RMSE for some quantiles are obtained as follows in Table IV.

TABLE IV. RMSE QRNN METHOD (HYBRID MODEL)

Quantile	RMSE
0.025	1.99
0.5	1.48
0.075	1.45

D. Comparing the Performance of Neural Network and Quantile Regression Neural Network

Based on RMSE and time series plot each method obtained the best method is QRNN. For CPI data can be drawn conclusion that is by using normalized transformation, the best method for CPI data is by using QRNN method with RMSE is 0.95 as in Table V and visualized by Fig. 7.

TABLE V. COMPARISON OF FORECASTING METHOD

Method	RMSE
NN	1.07
QRNN	0.95
Hybrid NN-QRNN	1.48

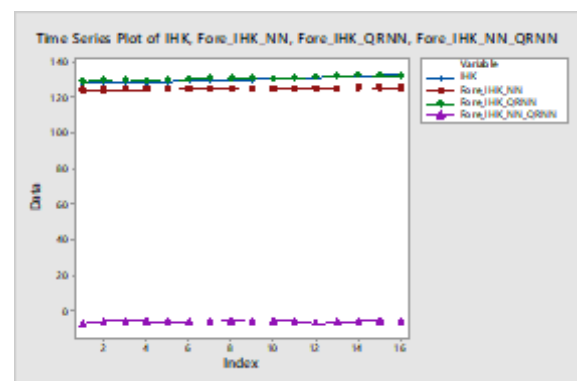


Fig. 7. Time Series Plot of Forecasting Method

CONCLUSION

Based on using NN and QRNN method on CPI data, can be drawn conclusion that is by using normalized transformation, the best method for CPI data is by using

QRNN method with comparing the value of RMSE. The results of RMSE for QRNN is 0.95.

SUGGESTION

Suggestion for further research on ARIMAX model, outlier detection is done to assume the normal distributed residual can be fulfilled. Then, do a strategy to avoid crossing the QRNN forecast hose.

REFERENCES

- [1] G. P. Zhang, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," *Neurocomputing*, vol. 20, pp. 159-175, 2003.
- [2] R. W. Koenker, & G. W. Bassett, *Quantile Regression*, Cambridge: Cambridge University Press, 2005.
- [3] Investopedia, Consumer Price Index, retrieved June 5th 2018 from <https://www.investopedia.com/terms/c/consumerpriceindex.asp>, 2018.
- [4] F. Kharimah, M. Usman, Widiarti, & F. Elfaki, "Time Series Modeling and Forecasting of the Consumer Price Index Bandar Lampung," *Sci. Int. (Lahore)*, vol. 27, pp. 4619-4624, 2015.
- [5] H. Norbert, A. Wanjoya, & A. Waititu, "Modeling and Forecasting Consumer Price Index (Case of Rwanda)," *American Journal of Theoretical and Applied Statistics*, vol. 5, pp. 101-107, 2016.
- [6] S. Makridakis, & M. Hibon, "The M3-Competition: Results, Conclusions, and Implications," *International Journal of Forecasting*, vol. 16, pp. 451-476.
- [7] J. W. Taylor, "A Quantile Regression Neural Network Approach to Estimating the Conditional Density of Multiperiod Returns," *Journal of Forecasting*, vol. 19, pp. 299-311, 2000.
- [8] Y. Fenf, A. Sudjianto, & Y. Zhang, "Robust Neural Network with Applications to Credit Portfolio Data Analysis," *Statistics and Its Interface*, vol. 3, pp. 437-444, 2010.
- [9] T. Terasvirta, C. Lin, & C. Granger, "Power of the Neural Network Linearity," *Journal of Time Series Analysis*, vol. 14, pp. 209-220.
- [10] A. J. Cannon, "Quantile Regression Neural Network: Implementation in R and Application to Precipitation Downscaling," *Computers and Geosciences*, vol. 37, pp. 1277-1284, 2011.
- [11] F.S. Crone & N. Kourentzes, "Input-Variable Specification for Neural Networks – an Analysis of Forecasting Low and High Time Series Frequency," *International Joint Conferences on Neural Networks*, pp. 619-626, 2009.