Quantile Regression Neural Network for Forecasting Inflow and Outflow in Yogyakarta

Farah Fajrina Amalia¹, Suhartono²*, Santi Puteri Rahayu³, and Novri Suhermi⁴
¹Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
²Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
³Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
⁴Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
suhartono@statistika.its.ac.id

Abstract: Quantile Regression Neural Network (QRNN) is a hybrid method that is developed based on quantile regression (QR) that can model data with non-homogeneous variance and neural network (NN) approach that can capture nonlinear patterns in the data. One example of real data that supposedly have such characteristics is the inflow and outflow of currency, where the inflow is the amount of money flow coming from banks and the public to Bank Indonesia (BI) while outflow is the flow of money out of BI to the banks and community. The data used in this research are inflow and outflow in Yogyakarta as many as 14 notes during January 2003 until December 2016 period. This study aims to forecast inflow and outflow in Yogyakarta using QRNN method and compare the results with individual method, i.e. ARIMAX-GARCH and NN. Based on RMSE and MdAE evaluation criteria, the results show that the best model for inflow of Rp100,000, Rp20,000 Rp10,000, and Rp5,000 are ARIMAX, while for the inflow Rp50,000, Rp2,000, Rp1,000, and outflow Rp100,000, Rp50,000, Rp20,000, Rp10,000, Rp5,000, Rp2,000, and Rp1,000 are QRNN. In general, the QRNN method yields better forecasting results and can capture well the effects of calendar variations on the data.

1. Introduction
One of the modern forecasting methods that is now widely studied is Neural Network (NN). The advantage of forecasting with NN can be done on nonlinear data. In addition, NN is a universal approach that has a high degree of accuracy without requiring a principle to meet [1]. Empirically, the data used for forecasting usually has a non-homogeneous variance. One method that can be used to model data with non-homogeneous variances is Quantile Regression (QR). The QR method has many advantages over the Ordinary Least Square (OLS) method. If the OLS needs an identical, independent, normal distributed residual, and no heteroscedasticity, the QR method does not require that assumptions to be fulfilled. Additionally, since the QR method estimates the quantile of conditional distribution, it is more resistant to the outliers of the OLS method.

The results, conclusions, and implications of M3-Competition indicated that combination of several different methods can yield in higher forecasting [3]. Taylor proposed a nonlinear nonparametric
method, namely Quantile Regression Neural Network (QRNN), which combined the advantages of quantile regression and neural network [4]. The other research used the QRNN method in the credit portfolio data showed that this method is more resistant to outliers when compared to linear regression and spline regression [5].

An example of Indonesia cases that can be modeled by QRNN is Bank Indonesia’s inflow and outflow data. There are two theoretical models that are usually used to calculate inflow and outflow by Bank Indonesia, i.e. the Error Correction Model and decomposition method [6,7].

Previous research about inflow and outflow forecasting showed that some methods that usually be used to overcome this problem are ARIMA, Time Series Regression, and ARIMAX methods. The result also indicated the presence of non-stationary pattern in the mean and variance [8]. Research on forecasting inflows and outflows in Bali showed that ARIMAX with the effects of calendar variations was the best model. In addition, it was known that the variance of data was not homogenous [9]. Previous research on the effects of calendar variation showed that the ARIMAX (with the effects of calendar variations) method yielded the highest accuracy especially for out-sample data, compared with Decomposition, ARIMA, and Neural Network [10]. Research on the application of Time Series Regression with Calendar Variation Effect on net flow data also showed that Eid al-Fitr significantly affected the currency net flow [11].

One of the conventional method that commonly used in forecasting is ARIMA. The development of the ARIMA model by adding an exogenous variable is often called the ARIMAX model. In this research, ARIMAX, NN and QRNN methods are used to forecast inflow and outflow data, then the accuracy of each method is compared by using MdAE evaluation criteria.

2. Literature Review

2.1. Autoregressive Integrated Moving Average (ARIMA)

ARIMA was first developed by George Box and Gwilym Jenkins for modeling time series data. ARIMA represents three models: autoregressive model (AR), moving average (MA), and autoregressive and moving average (ARMA) [12]. In general, the ARIMA model can be written as

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) a_t ,$$  \hfill (1)

where

- $\phi_p(B) =$ AR component of order $p$
- $\theta_q(B) =$ MA component of order $q$
- $(1-B)^d =$ operator of differencing order $d$
- $a_t =$ White Noise residual ($a_t \sim WN(0,\sigma^2_a)$).

The seasonal ARIMA model is

$$Y_t = \frac{\theta_q(B)\Theta_q(B^s)}{\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^d} a_t ,$$  \hfill (2)

The procedure stages of the Box-Jenkins ARIMA method are model identification, parameter estimation, diagnostic checks, and forecasting [13].

2.2. ARIMAX

ARIMAX model is an ARIMA model with the addition of $x$ variable. In this study, the $x$ variable used is the trend, seasonal, and effects of calendar variations. The ARIMAX model can be written as follows

$$Y_t = \beta_1 V_{1,t} + \beta_2 V_{2,t} + \ldots + \beta_p V_{p,t} + \frac{\theta(B)}{\phi_p(B)} a_t ,$$  \hfill (3)
where \( V_{p,j} \) is the dummy variable for the effect of \( p \)-th calendar variation. The ordering process \((p, q)\) is the same as the ARIMA model, which is done by identifying ACF and PACF plots on stationary data [13].

2.3. Neural Network

Neural network is a technique in machine learning that has been developed as a generalization of mathematical model from biological nervous system. The neural network model commonly used in forecasting is feed forward neural network [1]. An example of a forward neural network feed with three layers for forecasting time series data is shown in Figure 1.

![Feed forward neural network](image)

**Figure 1.** Feed forward neural network

The relationship between the output \( (Y_t) \) and the input \( (Y_{t-1}, Y_{t-2}, ..., Y_{t-p}) \) can be described as follows:

\[
Y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g(\beta_{jq}) + \sum_{i=1}^{p} \beta_i Y_{t-i} + a_j \tag{4}
\]

with \( \alpha_j (j = 0, 1, 2, ..., q) \), \( \beta_i (i = 0, 1, 2, ..., p; j = 1, 2, ..., q) \) is the model parameters or often referred as weights, \( p \) is the number of input nodes, and \( q \) is the number of hidden nodes. Equation (4) forms a nonlinear mapping of the past value \( (Y_{t-1}, Y_{t-2}, ..., Y_{t-p}) \) to the future value \( (Y_t) \) as follows:

\[
Y_t = f(Y_{t-1}, Y_{t-2}, ..., Y_{t-p}, w) + a_j \tag{5}
\]

with \( w \) is a vector of all parameters and \( f \) is a function determined from the network structure and weights. Thus, the neural network in Figure 1 is equivalent to the nonlinear autoregressive model [1].

2.4. Quantile Regression Neural Network

Quantile Regression Neural Network is a combination of Quantile Regression and Neural Network. For example, known \( \chi_i(t) \) predictors and \( y(t) \) responses. The output from the \( j \)-th hidden layer node is

\[
g_j(t) = \tanh \left( \sum_{i=1}^{I} \chi_i(t)w_{ij}^{(s)} + b_{ij}^{(s)} \right) \tag{6}
\]

with \( w_{ij}^{(s)} \) is the weight of the hidden layer whereas \( b_{ij}^{(s)} \) is the bias of the hidden layer [14]. The estimate of the conditional \( \hat{y}_\tau(t) \) on the \( \tau \)-th quantity is as follows

\[
\hat{y}_\tau(t) = f \left( \sum_{j=1}^{I} g_j(t)w_{ij}^{(o)} + b_{ij}^{(o)} \right) \tag{7}
\]

where \( w_{ij}^{(o)} \) is the weight of the output layer and \( b_{ij}^{(o)} \) is the bias of the output layer whereas \( f(\cdot) \) is a function of the output layer. The loss function is
\[
\rho^*(a(u)) = \begin{cases} 
    \tau h(u) & \text{if } u \geq 0 \\
    (\tau - 1) h(u) & \text{if } u < 0 
\end{cases}
\]

where \( u \) is the residual of the parameter estimate and \( h(u) \) is Huber norm with (9) i.e.

\[
h(u) = \begin{cases} 
    \frac{u^2}{2\epsilon} & \text{if } 0 \leq |u| \leq \epsilon \\
    |u| - \frac{\epsilon}{2} & \text{if } |u| > \epsilon 
\end{cases}
\]

where \( \epsilon \) is the threshold which value is determined. The error function to be optimized is

\[
E^*(a) = \frac{1}{N} \sum_{n=1}^{N} \rho^*(a) (y(t) - \hat{y}_n(t)).
\]

The optimization on the QRNN method uses a quasi-Newton algorithm to minimize \( E^*(a) \) with the following stages.

- Determine the relatively large initial value of \( \epsilon \)
- Obtain the values of weight and bias
- The process is repeated again using the values of weight and bias obtained and the value of \( \epsilon \) is getting smaller
- When the value of \( \epsilon \) is close to zero, the algorithm has converged and obtained the minimum \( E^* \) value.

2.5. Best Model Selection

Model evaluation needs to be done to determine the best forecasting model. The calculation method used is the Median Absolute Error (MdAE), a criterion based on residual [12] which is calculated by

\[
MdAE = \text{median}(|e_i|),
\]

where,

\[e_i = \hat{y}_{n+l} - \hat{y}_n(l)\]

\(\hat{y}_{n+l}\) = observation value of \( l \) step ahead

\(\hat{y}_n(l)\) = forecast value of \( l \) step ahead.

2.6. Cash Inflow and Outflow

Cash outflow represents the information concerning the flow of cash money issued by Bank Indonesia to banks and the public. Cash inflow represents the information concerning the flow of cash money coming from banks and the public to Bank Indonesia. Inflow and outflow became one of the considerations for Bank Indonesia in formulating the Money Distribution Plan [6].

3. Research Methodology

The data used in this research is secondary data from Bank Indonesia. Period of data is January 2003 to December 2016. Data in January 2003-December 2015 period are used as in-sample data while the data in January 2016-December 2016 are used as out sample data. In this study, the variables used are monthly inflow and outflow data in Yogyakarta in million IDR with details in Table I.

Predictors used are effects of calendar variations, trend patterns, and seasonality with the following details.

1) Dummy variables for the effects of calendar variations of Eid al-Fitr
   The effect of Eid al-Fitr is related to the week in the previous month (\( t-1 \)), in the month (\( t \)), and one month after (\( t+1 \)). Thus, the dummy variable used is as follows.
   a. Effect of one month before Eid al-Fitr
2nd International Conference on Statistics, Mathematics, Teaching, and Research
IOP Conf. Series: Journal of Physics: Conf. Series 1028 (2018) 012232
doi:10.1088/1742-6596/1028/1/012232

b. Effect of Eid al-Fitr month

\[ V_{j,t} = \begin{cases} 
1, & \text{month of Eid Fitr in } j\text{-th week} \\
0, & \text{others} 
\end{cases} \]

2) Variables for trend patterns are as follows.

The trend pattern is linear, then the variable used is \( t = 1, 2, \ldots, n \)

3) Dummy variables for seasonal patterns (months).

Since dummy regression will be used for the reconstruction of seasonal patterns, the dummy variables used are as follows.

\[ S_{1,t} = \begin{cases} 
1, & \text{January} \\
0, & \text{others} 
\end{cases} \]

\[ \vdots \]

\[ S_{12,t} = \begin{cases} 
1, & \text{December} \\
0, & \text{others} 
\end{cases} \]

| Table 1. Research variables |
|-----------------------------|
| **Variable** | **Data** | **Meaning** |
| \( Y_{1,t} \) | Rp100,000 currency in \( t \)-th period |
| \( Y_{2,t} \) | Rp50,000 currency in \( t \)-th period |
| \( Y_{3,t} \) | Rp20,000 currency in \( t \)-th period |
| \( Y_{4,t} \) | Rp10,000 currency in \( t \)-th period |
| \( Y_{5,t} \) | Rp5,000 currency in \( t \)-th period |
| \( Y_{6,t} \) | Rp2,000 currency in \( t \)-th period |
| \( Y_{7,t} \) | Rp1,000 currency in \( t \)-th period |
| \( Y_{8,t} \) | Rp100,000 currency in \( t \)-th period |
| \( Y_{9,t} \) | Rp50,000 currency in \( t \)-th period |
| \( Y_{10,t} \) | Rp20,000 currency in \( t \)-th period |
| \( Y_{11,t} \) | Rp10,000 currency in \( t \)-th period |
| \( Y_{12,t} \) | Rp5,000 currency in \( t \)-th period |
| \( Y_{13,t} \) | Rp2,000 currency in \( t \)-th period |
| \( Y_{14,t} \) | Rp1,000 currency in \( t \)-th period |

The steps of analysis performed in this research are as follows:

- Performing descriptive statistical analysis on inflow and outflow data in Yogyakarta Province
- Dividing the in-sample and out-sample data
- Modeling with ARIMAX
- Modeling with Neural Network
- Modeling with Quantile Regression Neural Network (QRNN) with 3 different inputs, namely:
  - Dummy variables and \( Y_t \) lags
  - TSR prediction and \( Y_t \) lags
  - TSR prediction and \( N_t \) (residual of TSR model) lags
- Comparing the ARIMAX-GARCH, NN, and QRNN methods based on the MdAE value to select the best forecasting method.
- Forecasting inflow and outflow with the best method.
4. Results

4.1. Cash Inflow and Outflow Characteristics
The time series plot of cash inflow and outflow in Yogyakarta Province is shown at Figure 2. It shows a fluctuating pattern, where the data from 2003 until the end of 2006 are increasing, then in 2007 until the end of 2010 there was no significant change, then in 2011 to 2016 there was a significant increase. In addition, the inflow and outflow data are influenced by Eid al-Fitr as shown by dashed lines in the figure.

![Figure 2. Time Series Plot of (a) Inflow and (b) Outflow](image)

4.2. Modeling of Inflow and Outflow using ARIMAX
The first step is to regress each of $Y_i$ variables with the trend, seasonal, and calendar variations. The next step is to identify the residuals of the time series regression model based on the ACF and PACF plots, to obtain the ARIMA model for each fraction. Based on the test results, it is known that the residuals of all outflow currencies have fulfilled the white noise assumption. However, there are currencies which do not have normal distributed residuals, i.e. Rp50,000, Rp20,000, Rp10,000, Rp5,000, Rp2,000 and Rp1,000. The next step is to perform the Lagrange Multiplier test, resulting that there is no ARCH/GARCH effect on all currencies. Figure 3 is the plot of the observed and forecasted data of the Rp100,000 inflow.

![Figure 3. The Results of ARIMAX at out-sample data](image)

Figure 3 shows that the observed data with the effects of calendar variations are beyond the interval limit which means that the ARIMAX model has not been able to capture the effects of calendar variations well.

4.3. Modeling of Inflow and Outflow using NN
In modeling using NN, it is important to determine the inputs to be used on the model. To capture trend patterns, seasonality, and effects of calendar variations, dummy variables are used in modeling with NN. In addition, the input used is the lags of the data. Based on the significant lags in PACF plot, the inputs for the neural network model of data inflow and outflow are shown in Table II.
To find out whether there is nonlinear relationship between the variables with the corresponding inputs, the Terasvirta test is performed and the result is all currencies have nonlinear relations with each lags. Neural network modeling uses 1 to 15 neurons to find the best model. Figure 4 is the plot of observed and forecasted data of Rp100,000 inflow.

4.4. Modeling of Inflow and Outflow using QRNN

4.4.1. QRNN with Dummy Variables and Y\_t Lags as Inputs
Using the same inputs from NN best model, the data are analyzed using the QRNN method. Figure 5 is plot of observed and forecasted data from a one stage QRNN model.

Based on Figure 5, there is a discrepancy in the interval forecast, in which the lower limit of the interval is higher than the upper limit of the interval, and the forecast point is beyond the interval limit. Therefore, it can be concluded that the forecast interval of the QRNN method is inconsistent and cannot be used.

4.4.2. QRNN with TSR Prediction dan Y\_t Lags as Inputs
The inputs used in this QRNN modeling are prediction values generated from the time series regression and Y\_t lags based on PACF plot. Then, comparison of observed and forecasted data of Rp100,000 inflow is shown in Figure 6.
4.4.3. QRNN with TSR Prediction and \( N_t \) Lags as Inputs

The inputs used in this QRNN modeling are prediction values generated from the time series regression and \( N_t \) lags based on PACF plot. Figure 7 is a plot of observed and forecasted data of Rp100,000 inflow.

![Figure 7. The Results of QRNN (TSR Prediction & \( N_t \) Lags)](image)

### 4.5. Comparison of ARIMAX, NN, and QRNN Models

Each result from modeling using ARIMAX, NN and QRNN is evaluated using MdAE. The MdAE ratio of the NN and QRNN methods against ARIMAX is used to determine the best model as shown in Table 3.

| Currency | ARIMAX | NN (One Stage) | QRNN (One Stage & Y_lags) | QRNN (CSR Fits & Y_lags) | QRNN (CSR Fits & Y_lags) |
|----------|--------|----------------|---------------------------|---------------------------|---------------------------|
| Rp100,000 | 1.00   | 3.85           | 1.94                      | 2.04                      | 1.75                      |
| Rp50,000  | 1.00   | 2.86           | 1.37                      | 0.95                      | 1.07                      |
| Rp20,000  | 1.00   | 2.42           | 1.95                      | 5.00                      | 1.00                      |
| Rp10,000  | 1.00   | 1.27           | 1.55                      | 2.05                      | 2.10                      |
| Rp5,000   | 1.00   | 1.27           | 1.54                      | 1.47                      | 2.29                      |
| Rp2,000   | 1.00   | 4.17           | 0.51                      | 0.55                      | 0.31                      |
| Rp1,000   | 1.00   | 0.38           | 0.28                      | 0.30                      | 0.25                      |
| Rp100,000 | 1.00   | 1.56           | 0.90                      | 0.49                      | 0.46                      |
| Rp50,000  | 1.00   | 1.56           | 1.12                      | 1.22                      | 0.75                      |
| Rp20,000  | 1.00   | 0.29           | 0.77                      | 0.57                      | 0.79                      |
| Rp10,000  | 1.00   | 0.31           | 0.51                      | 0.43                      | 0.31                      |
| Rp5,000   | 1.00   | 1.60           | 1.05                      | 0.50                      | 0.55                      |
| Rp2,000   | 1.00   | 1.01           | 0.71                      | 0.39                      | 0.21                      |
| Rp1,000   | 1.00   | 0.42           | 0.29                      | 0.25                      | 0.67                      |

Based on Table III, it is known that QRNN dominates as the best model for both inflow and outflow forecasting, i.e. 10 out of 14 currencies. The NN method has MdAE ratio greater than 1 in most currencies, indicating that the NN method does not produce better forecasts than ARIMAX methods. In Rp 1,000 outflow, the QRNN method with predictive TSR and lag \( Y_t \) inputs can reduce forecasting errors by 77%, compared to the ARIMAX model. However, there is crossing in the interval forecast so that the forecast interval of the QRNN method cannot be used. This may be due to insufficient number of observations in QRNN model training [14]. The lack of such observations can lead to crossings in QRNN's forecast intervals.
4.6. Inflow and Outflow Forecasting for Year 2017

Based on the best model, the forecast value of inflow and outflow for the year 2017 is shown in Table 4.

| Month       | Rp100,000 | Rp50,000 | Rp20,000 | Rp10,000 | Rp5,000 | Rp2,000 | Rp1,000 |
|-------------|-----------|----------|----------|----------|---------|---------|---------|
| January     | 996580    | 553970   | 20134    | 17637    | 11502   | 946     | 424     |
| February    | 66426     | 309279   | 14340    | 21630    | 9724    | 2138    | 422     |
| March       | 656916    | 408684   | 15998    | 14370    | 8026    | 2372    | 423     |
| April       | 649900    | 446518   | 13694    | 12286    | 6913    | 4752    | 421     |
| May         | 645867    | 42867    | 12177    | 10923    | 7177    | 2164    | 420     |
| June        | 674497    | 378502   | 10961    | 11882    | 5825    | 1509    | 419     |
| July        | 1156005   | 1057399  | 17906    | 20818    | 11371   | 2718    | 418     |
| August      | 869063    | 533045   | 17106    | 33636    | 29149   | 3456    | 860     |
| September   | 561641    | 453472   | 17468    | 20217    | 18979   | 3405    | 443     |
| October     | 709945    | 426983   | 16329    | 21111    | 16114   | 3404    | 418     |
| November    | 593219    | 595329   | 13070    | 12467    | 9304    | 3366    | 419     |
| December    | 548215    | 586727   | 13903    | 10563    | 9139    | 3404    | 418     |

5. Conclusion

Based on the analysis and discussion that has been done, it can be concluded that the inflow and outflow of currency in Yogyakarta is influenced by the trend, seasonal, and effects of calendar variation Eid al-Fitr. The currencies Rp50,000, Rp2,000, and Rp1,000 inflow, and Rp100,000 to Rp1000 outflow are better modeled by QRNN. Hence, QRNN is more accurate than ARIMAX and NN methods for forecasting most of the inflow and outflow currencies. Moreover, a strategy should be taken to avoid crossing in the QRNN method interval forecast, such as bootstrap or parameter regularization.

References

[1] G. P. Zhang. 2003 Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 20 159-175.
[2] R. Koenker. 2005 Quantile Regression. Cambridge: Cambridge University Press
[3] S. Makridakis, M. Hibon. 2000 The M3-Competition: results, conclusions, and implications. International Journal of Forecasting, 16 451-476.
[4] J. W. Taylor. 2000 A Quantile Regression Neural Network Approach to Estimating the Conditional Density of Multiperiod Returns. Journal of Forecasting, 19 299-311.
[5] Y. Feng, A. Sudijanto, Y. Zhang. 2010 Robust neural network with applications to credit portfolio data analysis. Statistics and Its Interface, 3 437-444.
[6] H. Sigalingging, E. Setiawan, H. D. Sihaloho. 2004 Kebijakan Pengedaran Uang Di Indonesia Jakarta: Pusat Pendidikan dan Studi Kebanksentralan
[7] E. Yulianto. 2011 Yuk Menghitung Kebutuhan Uang. Gerai Info: Newsletter Bank Indonesia
[8] N. I. Rachmawati, Setiawan, Suhartono. 2015 Peramalan Inflow dan Outflow Uang Kartal Bank Indonesia di Wilayah Jawa Tengah dengan Menggunakan Metode ARIMA, Time Series Regression, dan ARIMAX. Jurnal Sains dan Seni ITS, 42 323-329.
[9] I. M. G. M. Dana. 2016 Peramalan Inflow dan Outflow Sebagai Proyeksi Kebutuhan Uang
Kartal di Provinsi Bali Menggunakan Fungsi Transfer, TBATS, ARIMA, dan ARIMAX. Undergraduate Thesis, Institut Teknologi Sepuluh Nopember, Surabaya

[10] Suhartono, M. H. Lee, N. A. Hamzah. 2010 Calendar variation model based on Time Series Regression for sales forecasts: The Ramadhan effects. Proceeding of Regional Conference on Statistical Sciences 30-41.

[11] R. E. Wulansari, E. Suryanto, K. Ferawati, Suhartono. 2014 Penerapan Time Series Regression with Calendar Variation Effect pada Data Netflow Uang Kartal Bank Indonesia Sebagai Solusi Kontrol Likuiditas Perbankan di Indonesia. Statistika: Forum Teori dan Aplikasi Statistika, 142 59-68.

[12] W. W. S. Wei. 2006. Time Series Analysis : Univariate and Multivariate Methods. Boston : Pearson Addison Wesley

[13] M. H. Lee, Suhartono, N. A. Hamzah. 2010 Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect. Proceeding of Regional Conference on Statistical Sciences 349-361.

[14] A. J. Cannon. 2011 Quantile regression neural networks: Implementation in R and application to precipitation downscaling. Computers and Geosciences, 37, 1277-1284.

[15] J. S. Armstrong, F. Collopy. 1992 Error Measures For Generalizing About Forecasting Methods: Empirical Comparison. International Journal of Forecasting 8 69-80.