Hybrid ARIMAX quantile regression method for forecasting short term electricity consumption in east java

M Prastuti\(^1\), Suhartono\(^2\), and NA Salehah\(^2\)

\(^1\) Business Statistics Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia 60111
\(^2\) Statistics Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia 60111

Email: suhartono@statistika.its.ac.id

Abstract. The need for energy supply, especially for electricity in Indonesia has been increasing in the last past years. Furthermore, the high electricity usage by people at different times leads to the occurrence of heteroscedasticity issue. Estimate the electricity supply that could fulfilled the community’s need is very important, but the heteroscedasticity issue often made electricity forecasting hard to be done. An accurate forecast of electricity consumptions is one of the key challenges for energy provider to make better resources and service planning and also take control actions in order to balance the electricity supply and demand for community. In this paper, hybrid ARIMAX Quantile Regression (ARIMAX-QR) approach was proposed to predict the short-term electricity consumption in East Java. This method will also be compared to time series regression using RMSE, MAPE, and MdAPE criteria. The data used in this research was the electricity consumption per half-an-hour data during the period of September 2015 to April 2016. The results show that the proposed approach can be a competitive alternative to forecast short-term electricity in East Java. ARIMAX-QR using lag values and dummy variables as predictors yield more accurate prediction in both in-sample and out-sample data. Moreover, both time series regression and ARIMAX-QR methods with addition of lag values as predictor could capture accurately the patterns in the data. Hence, it produces better predictions compared to the models that not use additional lag variables.

1. Introduction
Quantile regression is an estimation method that can overcome the data distribution that is not uniform. Data that has a non-uniform pattern is called heteroscedasticity, or a non-fulfillment of homoscedasticity assumption on regression analysis using Ordinary Least Square (OLS) method is also often called heteroscedasticity. This problem can be overcome by using the quantile regression method. The parameter estimation in the OLS method only provide a solution of the mean problem, so Koenker and Basset developed an alternative method of quantile regression [1]. Quantile regression extends the calculation of coefficient values across various quantiles, in order to provide more complete picture of the data conditions. This method is robust against outlier data, so it is highly recommended to be used to analyze a number of data that is not symmetrical and has a not homogeneous distribution. Forecasting using a quantile regression approach will produce better accuracy than the classical conditional mean method [2].

Combining two or more methods or hybrid method tend to improve the accuracy of prediction [3], so many researchers are using and developing it now. Hybrid method has been widely used in various
field of science. For examples the development of combined method of quantile-based Regression Neural Network (QRNN) has been used to estimate the conditional density of multiperiod return value and compared with the results of the estimation using the GARCH-based Quantile Regression, and it show that the QRNN method is able to predict conditional density well [4]. Other studies apply hybrid ARIMA-LR (Linear Regression) method to predict the arrival of patients in an ER in China [5]. In 2015, Arunraj and Ahrens proposed hybrid SARIMAX-Quantile Regression model then applied it to forecast daily food sales and this study shows that hybrid SARIMAX-QR model provides better forecast accuracy for out-sample data than other individual method and provides better interval prediction [6].

This research will conduct a study about the development of hybrid models for forecasting the electricity energy consumption data in East Java. The data is suspected to have a calendar variation effect, resulting in the presence of certain patterns such as seasonal patterns, holiday effects, and effects of special days. This cause indicates that there were heteroscedasticity case in the electricity energy consumption data, so generating the forecast value of its data will not be easy. Based on this condition, this study aims to develop hybrid method using hybrid ARIMAX Quantile Regression (ARIMAX-QR) method for forecasting of electric energy consumption data in East Java. This method is expected to be able to yield forecasting model for data of electric energy consumption in East Java properly and minimize the cause of any forecasting errors. Forecasting of data on electricity consumption in East Java is very important, because it can be used as a material for distribution of electricity plan more efficient, so the distribution of electricity energy to consumers in East Java can be done optimally and fulfills the needs.

2. Literature Review

2.1. ARIMA

ARIMA is a time series analysis method composed of two models, namely autoregressive model (AR) and Moving Average (MA). ARMA model is composed of AR order $p$ and MA order $q$, while ARIMA ($p$, $d$, $q$) is ARMA model ($p$, $q$) which get differencing as much as $d$. In the short-run electricity consumption data, double seasonal patterns are present in seasonal per-half hour and seasonal daily so it called has double seasonal multiplicative or Double Seasonal ARIMA patterns [7]. The Double Seasonal ARIMA model is written with ARIMA ($p,d,q$)($P_1,D_1,Q_1$)$^s$($P_2,D_2,Q_2$)$^d$ which has the following common form,

$$\phi(B)\Phi(B^s)\phi(B^d)(1-B)^d(1-B^s)^D(1-B^d)^D\gamma Y_t = \theta(B)\Theta(B^s)\theta(B^d)a_t.$$  \hspace{1cm} (1)

In performing ARIMA model, there are several steps that must be done, such as model identification, parameter estimation and testing, model diagnostic, model selection, and forecasting [8].

2.2. Time Series Regression

The models in time series regression for the electricity forecasting are involved by trend, seasonal patterns, special days, and autocorrelated errors [7]. The effect of trend on time series regression is expressed by $t$. Weekly seasonal pattern consists of 7 variables i.e. Monday ($H_{1t}$), Tuesday ($H_{2t}$), Wednesday ($H_{3t}$), Thursday ($H_{4t}$), Friday ($H_{5t}$), Saturday ($H_{6t}$), and Sunday ($H_{7t}$). The daily seasonal pattern consists of 48 variables because this study uses the electricity consumption data per half an hour ($D_{1t}, D_{2t}, \ldots, D_{48t}$). Otherwise, the pattern of special day is indicated by dummy variable, $L_{jt}$. The model of the time series regression used in this study is,

$$Y_t = \eta + \gamma_1 H_{1t} + \gamma_2 H_{2t} + \ldots + \gamma_7 H_{7t} + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \ldots + \alpha_{48} D_{48t} + \omega_1 L_{1t} + \omega_2 L_{2t} + \ldots + \omega_7 L_{7t} + \epsilon_t$$  \hspace{1cm} (2)

From the time series regression equation above there are 64 predictors of dummy variables. If the error does not fulfill the white noise assumption then lag of series will be used as an additional
predictor variable. The selection of lag can be determined based on ACF and PACF plots [9]. Here are the steps of model building in the time series regression.

a. Parameter Estimation
Parameter estimation using Ordinary Least Square (OLS) minimizes the sum of squares of deviations or errors of observation values. The equation of a simple linear regression model is as follows [10]:

$$ Y = X\beta + \varepsilon, $$

with $\beta$ being a vector containing the parameters that will be estimated from the model. Equation 3 then differentiated to each parameter and the estimated parameter that obtained from this process can be written as,

$$ \hat{\beta} = (X'X)^{-1}X'Y, $$

where,

$$ Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & H_{11} & L \\ 2 & H_{21} & L \\ \vdots & M & L \\ n & H_{n1} & L \end{bmatrix}, \quad \beta = \begin{bmatrix} \eta \\ M \\ \omega \end{bmatrix}. $$

b. Parameter Significance Test
Parameter significance test is done partially. Partial testing is a test that can be done as many parameters in the regression model [10]. The hypothesis testing parameters can be written as,

$$ H_0 : \beta_j = 0 $$

$$ H_1 : \beta_j \neq 0, \quad j = 1, 2, ..., k. $$

The $t$ statistic will be used to test the significance of the parameters,

$$ t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}. $$

Reject $H_0$ if $|t| > t_{\alpha/2,n-k}$ or if p-value less than $\alpha$, with $n$ is the sample size and $k$ is the number of regression coefficients.

2.3. Quantile Regression
Quantile is a technique of dividing a group of data into several equal parts, after the data is sorted from the smallest or the largest [11]. Quantile regression estimation is more robust to the outlier(s), and heteroscedasticity. It is also does not need any assumption testing. Heteroscedasticity or unequal variation implies that there is more than one slope representing the relationship between response and predictor measured on a subset of these factors. Quantile regression estimates the slope(s) from minimum to maximum response, giving a more complete explanation of the relationship between variables that are sometimes missed by other regressions [12].

The quantile regression approach assumes the various quantile functions of a $Y$ distribution as a function of $X$. The use of this regression method is performed by divide or separate the data into several groups that suspected to have different alleged values on the quantile. The estimation of the response on the quantile regression uses the value of quantile-$\tau$ of $Y$ with certain $X$. The model of a quantile regression equation can be written as,

$$ y = Q_{\tau}(Y | X) = X\beta(\tau) + u, $$

where,
\[
\begin{align*}
\mathbf{y} &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \\
\mathbf{X} &= \begin{bmatrix} x_{11} & x_{21} & K & x_{p1} \\ x_{12} & x_{22} & K & x_{p2} \\ \vdots & \vdots & M & \vdots \\ x_{1n} & x_{2n} & K & x_{pn} \end{bmatrix}, \\
\mathbf{\beta}(\tau) &= \begin{bmatrix} \beta_1(\tau) \\ \beta_2(\tau) \\ \vdots \\ \beta_p(\tau) \end{bmatrix}, \\
\mathbf{u} &= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}
\end{align*}
\]

and \(0 < \tau < 1\).

Quantile regression weighted \(\tau\) for positive error and \((1-\tau)\) for negative error so the parameter estimation on quantile regression [11] can be expressed in the following form,

\[
\mathbf{\hat{\beta}}(\tau) = \min_{\mathbf{\beta} \in \mathbb{R}^p} \left[ \sum_{y \in \{y \leq \mathbf{X}\mathbf{\beta}\}} \tau |y - \mathbf{X}\mathbf{\beta}(\tau)| + \sum_{y \in \{y > \mathbf{X}\mathbf{\beta}\}} (1-\tau) |y - \mathbf{X}\mathbf{\beta}(\tau)| \right].
\]

where \(\mathbf{\hat{\beta}}(\tau)\) is a regression coefficient on quantile-\(\tau\).

### 2.4. ARIMAX Quantile Regression

This method combined some univariate methods with quantile regression approach. The forecast values are done against quantile 0.50 or often called as median estimate. There are two models of hybrid ARIMAX Quantile Regression that used in this research.

**2.4.1. ARIMAX Quantile Regression Model 1**

In this ARIMAX-QR modelling its uses the dummy variable and lag of the data as predictor variable in quantile regression. ACF and PACF plot from stationary error can be used to determine the lag. Thus, ARIMAX-QR model with additional lag variables could be written as follows:

\[
\hat{Y}_i(\tau) = Q_{\tau}(Y_i | t, H_{1,t}, ..., H_{7,t}, D_{1,t}, ..., D_{9,t}, L_{4,t}, ..., L_{9,t}, Y_{1-t}, ..., Y_{p-t}),
\]

where \(t, H_{1,t}, ..., H_{7,t}, D_{1,t}, ..., D_{9,t}, L_{4,t}, ..., L_{9,t}\) are exogenous variable(s).

**2.4.2. ARIMAX Quantile Regression Model 2**

This model is fitted by using its predicted values using the ARIMA method of data as well as other exogenous variable which are dummy variable(s) as predictor for quantile regression. This model can be written as follows:

\[
\hat{Y}_i(\tau) = Q_{\tau}(Y_i | \hat{Z}_i, t, H_{1,t}, ..., H_{7,t}, D_{1,t}, ..., D_{9,t}, L_{4,t}, ..., L_{9,t}),
\]

where \(\hat{Z}_i\) is prediction value of \(Y_i\) that developed using ARIMA model.

### 2.5. Model Evaluation

One of the purposes of forecasting is to improve the accuracy of forecast results and minimize the forecast error. This research using cross-validation approach to evaluate the model, using the model goodness criteria based on Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Median Absolute Percentage Error (MdAPE), with the following formula [13],

\[
\text{RMSE} = \sqrt{\frac{1}{M} \sum_{t=1}^{M} e_t^2},
\]

\[
\text{MAPE} = \frac{1}{M} \sum_{t=1}^{M} \left| \frac{e_t}{Y_{n+t}} \right| \times 100,
\]

\[
\text{MdAPE} = \text{median} \left( \left| \frac{e_t}{Y_{n+t}} \right| \right) \times 100,
\]

where \(e_t = Y_{n+t} - \hat{Y}_t(t)\).
3. Methodology

3.1. Data and Variable

The data used in this research was electricity energy consumption per half hour data in East Java. The data was secondary data which obtained from the subsidiary of PT. PLN (Persero) is PLN P2B (Distribution and Control Load Center) in East Java region. Period of this data was from September 1, 2015 to April 30, 2016. In analyzing this research, the data was divided into two parts: in-sample data used for modeling, and out-sample data to validate the method used. In-sample data include data from 01 September 2015 to 29 February 2016 in period, and out-sample includes data from 1 March 2016 to 30 April 2016.

The dummy variables used in this study include trend dummy, per-half hour, days and special days (which occur other than on Sunday), and can be presented in Table 1.

Table 1 Dummy Variables

| Dummy Variable | Explanation |
|----------------|-------------|
| Trend          | $t = 1, 2, 3, \ldots, 11664$ |
| Days           | $H_{1,t} = \begin{cases} 1, & \text{Monday} \\ 0, & \text{others} \end{cases}$ |
|                 | $H_{2,t} = \begin{cases} 1, & \text{Tuesday} \\ 0, & \text{others} \end{cases}$ |
|                 | $M \quad H_{7,t} = \begin{cases} 1, & \text{Sunday} \\ 0, & \text{others} \end{cases}$ |
| Per-half Hour  | $D_{1,t} = \begin{cases} 1, & \text{at 00.30} \\ 0, & \text{other} \end{cases}$ |
|                 | $D_{2,t} = \begin{cases} 1, & \text{at 01.00} \\ 0, & \text{other} \end{cases}$ |
|                 | $M \quad D_{4,t} = \begin{cases} 1, & \text{at 24.00} \\ 0, & \text{other} \end{cases}$ |
| Special Days   | $L_{1,t} = \begin{cases} 1, & \text{special day} \ i \\ 0, & \text{other days} \end{cases}$ |

Special days used as dummy variables ($L_{i,t}$) are special days or national holidays occurring in the period of observation of September 2015 to April 2016 which occur other than Sunday. The list of special days used in this research is presented in Table 2.

Table 2 Special Days List

| Dummy Variable | Special Day               | Date             |
|----------------|---------------------------|------------------|
| $L_{1,t}$      | Feast of Sacrifice (Eid Adha) | 24 September 2015 |
| $L_{2,t}$      | Islamic New Year           | 14 October 2015  |
| $L_{3,t}$      | Birth of Prophet (Mawlid)  | 24 December 2015 |
| $L_{4,t}$      | Christmas                  | 25 December 2015 |
| $L_{5,t}$      | New Year’s Day             | 01 January 2016  |
| $L_{6,t}$      | Chinese New Year           | 02 February 2016 |
| $L_{7,t}$      | Nyepi (Hindu’s Holly Day)  | 09 March 2016    |
| $L_{8,t}$      | Good Friday                | 25 March 2016    |
4. Results and Analysis

4.1. Exploration of Electricity Consumption Characteristics
The amount of electricity consumption was influenced by the needs of each customer. The diverse customer activity that causes electricity consumption in East Java is fluctuating. Based on data on electricity consumption used from September 1, 2015 until September 30, 2016 consists of 19,008 observations. The average electricity consumption in East Java is 4,064.7 MW with standard deviation of 484.9 MW. Electricity consumption in East Java is the lowest at 07.00 with electricity consumption of 3,530.0 MW. This happens because at 7:00 am many people are just leaving for work so that not many people who use electronic equipment such as television, lights, and computers. This causes the consumption of electricity is still a little not too high, while the highest consumption occurred at 18.30 with a consumption of 4,743.2 MW. It was allegedly because the public had returned home after work and used various electronic equipment so that the use of electricity consumption reached the highest consumption.

The volatility of electricity consumption was not only influenced by hours of use but also affected by day. The highest average electricity consumption occurred on Tuesday at 4,029.8 MW. Generally, electricity consumption on weekdays is higher than on weekend (Saturday and Sunday). This happens because on Saturday and Sunday offices, factories or government agencies do not use electricity as much as on weekdays, such as turning on production machines in factories, turning on computers, lights, air conditioners and so on. The lowest electricity consumption occurred on Sunday with an average of 3,639.9 MW. This is supposedly happened because on weekend (Saturday and Sunday) most people often spend time for recreation outside home, and enjoying their break time, so the electricity consumption not as much as on weekdays.

![Figure 1. Time Series (a) and Mean Plot (b) of Electricity Consumption in East Java](image)

Time series plot of electricity consumption in East Java Region shown in Figure 1. It is shown that from September 1, 2015 to April 30, 2016 the electricity consumption was increase. The electricity consumption on the weekdays from Monday to Thursday is relatively the same as shown in Figure 1b. But on Monday, from 0:30 to 05:00 and Friday at 12:00 to 12:30 shows the pattern of electricity consumption tend to be lower than the other weekdays in the same hour. On Monday at 0.30 to 05.00, electricity consumption is low, then it starts to increase significantly at 07.00 because most people started their activities. Otherwise, electricity consumption on Friday at 12.00 to 12.30 is lower than other working days because most people in East Java are doing the Friday prayer service.

4.2. Time Series Regression
Forecasting electricity consumption using time series regression was done by regressing the electricity consumption which is the response with the predictor such as dummy trend, hours, days, and special days (non-weekly national holidays). The model formed is shown in Table 3. From this model it is found that the residual still does not fulfill the white noise assumption, allegedly due to the effects of autocorrelated error. Therefore, modeling the data with adding lag of the series as a predictor was also done.
The ACF pattern of the data indicates a peak at lags 48, 96, 144, ... which indicates the daily seasonal period, and the peak in lag 336, 672, 1008, ... which denotes a seasonal period of the week. The PACF pattern shows a slow down to zero patterns. So it is concluded that the data has not been stationary. The process of differencing three times lag 1, 48, and 336 was done to make the data becomes stationary. From this result, lags used as predictors are lag 1, 48, and 336 (Y_t−1, Y_t−48, Y_t−336), where these lags show the seasonal patterns present in the data. In addition to these lags, also used its multiplicative lags such as lag 49, 337, 384 and 385 (Y_t−49, Y_t−337, Y_t−384, Y_t−385).

### Table 3 Time Series Regression Model for Electricity Consumption in East Java

| Predictor(s) | Model |
|--------------|-------|
| Dummy        | \( \hat{Y}_t = -0.009t + 3909H_{1,t} + 3998H_{2,t} + 3972H_{3,t} + 3994H_{4,t} + 4010H_{5,t} + 3759H_{6,t} + 3428H_{7,t} - 65.2D_{1,t} - 104.7D_{2,t} - 149.0D_{3,t} - 185.4D_{4,t} - 221.6D_{5,t} - 245.1D_{6,t} - 236.4D_{7,t} - 167.6D_{8,t} - 54.5D_{9,t} + 17.8D_{10,t} - 46.6D_{11,t} - 177.0D_{12,t} - 281.4D_{13,t} - 305.6D_{14,t} - 148.5D_{15,t} - 15.9D_{16,t} + 134.9D_{17,t} + 195.5D_{18,t} + 252.2D_{19,t} + 315.4D_{20,t} + 358.5D_{21,t} + 371.9D_{22,t} + 314.7D_{23,t} + 188.9D_{24,t} + 185.4D_{25,t} + 302.9D_{26,t} + 415.2D_{27,t} + 403.1D_{28,t} + 371.1D_{29,t} + 325.2D_{30,t} + 315.6D_{31,t} + 284.8D_{32,t} + 294.9D_{33,t} + 376.8D_{34,t} + 648.7D_{35,t} + 855.3D_{36,t} + 893.6D_{37,t} + 882.7D_{38,t} + 881.4D_{39,t} + 849.7D_{40,t} + 777.4D_{41,t} + 655.8D_{42,t} + 477.7D_{43,t} + 323.0D_{44,t} + 254.1D_{45,t} + 150.6D_{46,t} + 73.6D_{47,t} - 882.8L_{4,t} - 414.2L_{5,t} - 609.6L_{6,t} - 764.1L_{7,t} - 1078L_{8,t} - 703.7L_{9,t} |
| Dummy + Lag  | \( \hat{Y}_t = 0.00t + 58.9H_{1,t} + 52.5H_{2,t} + 51.0H_{3,t} + 53.9H_{4,t} + 52.9H_{5,t} + 44.1H_{6,t} + 37.9H_{7,t} + 1.9D_{1,t} + 11.5D_{2,t} + 8.1D_{3,t} + 9.3D_{4,t} + 10.1D_{5,t} + 13.5D_{6,t} + 25.3D_{7,t} + 49.0D_{8,t} + 66.8D_{9,t} + 52.7D_{10,t} + 2.8D_{11,t} - 23.1D_{12,t} - 15.7D_{13,t} + 13.4D_{14,t} + 80.7D_{15,t} + 74.9D_{16,t} + 84.8D_{17,t} + 52.8D_{18,t} + 52.3D_{19,t} + 56.6D_{20,t} + 49.8D_{21,t} + 39.2D_{22,t} + 12.6D_{23,t} - 15.7D_{24,t} + 29.5D_{25,t} + 75.6D_{26,t} + 75.9D_{27,t} + 30.1D_{28,t} + 22.2D_{29,t} + 17.9D_{30,t} + 28.8D_{31,t} + 21.3D_{32,t} + 37.6D_{33,t} + 63.7D_{34,t} + 133.6D_{35,t} + 116.6D_{36,t} + 59.8D_{37,t} + 40.4D_{38,t} + 43.7D_{39,t} + 31.6D_{40,t} + 16.7D_{41,t} - 4.91D_{42,t} - 28.6D_{43,t} - 23.3D_{44,t} + 6.7D_{45,t} - 7.1D_{46,t} + 0.7D_{47,t} - 39.4L_{1,t} - 19.7L_{2,t} - 29.1L_{3,t} - 22.6L_{4,t} - 30.7L_{5,t} - 32.8L_{6,t} + 0.91Y_{t-1} + 0.2Y_{t-48} - 0.2Y_{t-49} + 0.4Y_{t-336} - 0.4Y_{t-337} - 0.02Y_{t-384} - 0.03Y_{t-385} |

Both models were used to predict the out-sample data to check its performance in forecasting the several periods ahead data. These forecast values will be compared with its actual data (out-sample). The out-sample forecast plot of each model formed is shown in Figure 2. It is shown that the forecast values obtained by time series regression model were good enough to capture the electricity consumption data. After adding lag of the data as a predictor (Figure 2b), the forecast values tend to be better to capture the pattern of fluctuations than model that do not use lag as predictor (Figure 2a).

![Figure 2. Out-sample Forecast of Time Series Regression Model](image-url)
4.3. ARIMAX Quantile Regression

As we know, there are two models of ARIMAX Quantile Regression (ARIMAX-QR) applied in this research. ARIMAX-QR model 1 use lag and dummy variables as predictors. Determination of lag that used was same as the time series regression model using lag (subsection 4.2), lag 1, 48, and 336 ($Y_{t-1}, Y_{t-48}, Y_{t-336}$) and its multiplicative lags such as lag 49, 337, 384 and 385 ($Y_{t-49}, Y_{t-337}, Y_{t-384}, Y_{t-385}$) will be used as predictors. The last ARIMAX-QR model is model that used dummy variables and predicted values of ARIMA model for the data. Electricity consumption data contains two seasonal patterns, which are per-half hour and weekly patterns so Double Seasonal ARIMA need to be formed. For convenience, this research use ARIMA(0,1,1)(0,1,1)_{48}^6(0,1,1)_{336}^6 model as $\hat{Z}_t$. The model formed can be shown in Table 4.

| ARIMAX-QR | Model |
|-----------|-------|
| Model 1   | $\hat{Y}_t = 0.00r + 34.9H_{17} + 28.9H_{27} + 29.9H_{37} + 31.0H_{47} + 29.7H_{57} + 23.3H_{67} + 19.6H_{77} + 3.5D_{17} + 12.4D_{27} + 13.7D_{37} + 12.9D_{47} + 15.4D_{57} + 19.5D_{67} + 34.2D_{77} + 54.2D_{87} + 76.4D_{97} + 53.2D_{107} + 2.1D_{117} - 20.7D_{127} - 16.3D_{137} + 20.6D_{147} + 99.2D_{157} + 88.1D_{167} + 98.1D_{177} + 58.9D_{187} + 55.7D_{197} + 64.4D_{207} + 53.8D_{217} + 40.4D_{227} + 19.7D_{237} - 18.0D_{247} + 34.1D_{257} + 83.1D_{267} + 88.8D_{277} + 29.9D_{287} + 23.3D_{297} + 21.3D_{307} + 30.1D_{317} + 25.9D_{327} + 36.5D_{337} + 66.5D_{347} + 139.2D_{357} + 130.3D_{367} + 52.1D_{377} + 40.5D_{387} + 45.5D_{397} + 30.6D_{407} + 14.2D_{417} - 7.3D_{427} - 29.2D_{437} - 26.4D_{447} + 7.6D_{457} - 6.5D_{467} - 1.9D_{477} - 28.2L_{47} - 24.4L_{48} - 25.4L_{49} - 15.7L_{50} - 25.4L_{51} - 24.9L_{52} + 0.9Y_{t-1} + 0.2Y_{t-48} - 0.2Y_{t-337} - 0.4Y_{t-384} - 0.02Y_{t-185} |
| Model 2   | $\hat{Y}_t = 0.00r + 83.6H_{17} + 85.3H_{27} + 88.3H_{37} + 86.5H_{47} + 86.0H_{57} + 80.8H_{67} + 72.4H_{77} + 3.0D_{17} + 4.5D_{27} - 1.9D_{37} - 3.1D_{47} - 1.2D_{57} + 0.5D_{67} - 3.5D_{77} + 3.2D_{87} + 2.3D_{97} + 3.1D_{107} + 5.7D_{117} - 0.6D_{127} - 8.7D_{137} + 3.1D_{147} + 1.9D_{157} + 4.7D_{167} + 4.0D_{177} + 9.6D_{187} + 11.6D_{197} + 7.9D_{207} + 15.1D_{217} + 8.8D_{227} + 12.8D_{237} + 8.7D_{247} + 1.5D_{257} + 11.5D_{267} + 7.6D_{277} + 14.1D_{287} + 20.6D_{297} + 4.5D_{307} + 11.2D_{317} + 7.7D_{327} + 0.9D_{337} + 13.8D_{347} + 9.7D_{357} + 25.6D_{367} + 11.9D_{377} + 24.2D_{387} + 19.7D_{397} + 23.1D_{407} + 11.7D_{417} + 21.2D_{427} + 14.2D_{437} + 12.1D_{447} + 12.0D_{457} + 4.7D_{467} + 1.3D_{477} - 16.4L_{47} - 19.9L_{48} - 16.9L_{49} - 16.3L_{50} - 18.7L_{51} - 14.2L_{52} + 0.9Y_{t-384} |

From each ARIMAX-QR model formed, it will be used to forecast the out-sample data. The result of this process then compared to the actual as shown in Figure 3.

**Figure 3.** Out-sample Forecast of ARIMAX Quantile Regression Model

The predicted values generated by ARIMAX-QR model 1 (Figure 3a) show good results, where the forecast values tend to capture patterns from the data. The use of ARIMA prediction results in ARIMAX-QR model 2 aims to handle patterns that could not be captured using the decomposition model. The ARIMAX-QR model 2 prediction results as shown in Figure 3b show the disadvantages of...
ARIMA method, where ARIMA method are suitable only for short-term forecasting. For long-term forecasting, ARIMA tends to produce a prediction that is getting wider or farther from actual data. Hence, this model shows that the longer the period to be predicted then the results will tend to farther from the actual data. It implies model 2 yields more accurate for short-term forecasting.

4.4. Comparison between Time Series Regression and ARIMAX Quantile Regression

Based on the results at the previous chapter about the forecasting of electricity consumption in East Java using Time Series Regression (TSR) and ARIMAX Quantile Regression (ARIMAX-QR) method, Table 5 shows goodness values at out-sample data, i.e. RMSE, MAPE and MdAPE for each model.

| Method          | RMSE   | MAPE   | MdAPE  |
|-----------------|--------|--------|--------|
| TSR             |        |        |        |
| Dummy           | 220.787| 4.573  | 4.397  |
| Dummy + Lag     | 189.935| 4.474  | 4.256  |

| ARIMAX-QR       |        |        |        |
| Model 1         | 177.982| 3.374  | 2.761  |
| Model 2         | 1494.162| 29.979 | 27.833 |

Note: The bold values show the best model in each method

The best model formed from modelling electricity consumption data using time series regression (TSR) method is a model involving lag and dummy variables as predictors compared to models involving only dummy variables (Table 5). For modeling using hybrid ARIMAX Quantile Regression (ARIMAX-QR), the best selected model is model 1 which use dummy and lag as predictors in the quantile regression model compared to the other model. When comparing the best TSR model with the best ARIMAX-QR model it was found that the ARIMAX-QR model yields a prediction value with a smaller error rate than other models. To find out the predicted performance of the best TSR model and the best ARIMAX-QR model, an adaptive approach for RMSE, MAPE and MdAPE values of each model was used. The adaptive RMSE, MAPE and MdAPE plots of each model for 7-days ahead forecast period are shown in Figure 4.

![Figure 4. Adaptive Accuracy of Best Model](image)

It is shown that the accuracy of the ARIMAX Quantile Regression model is better than Time Series Regression model (Figure 4). This is indicated by a lower forecast error value than Time Series Regression model, for forecasting the next 1 day, 2 days ahead, and still low until the next 7 days forecast. PT PLN (Persero) set the benchmark maximum forecast error of 2%. Both models produce an adaptive MAPE and MdAPE less than the benchmark set by the PLN, which means both models are well used to forecast data on short-term power consumption in East Java. Based on MAPE criteria, it is also known that Time Series Regression model is good to forecast electricity load in East Java up to 6 days ahead. Otherwise, it will cause prediction up to 7 days ahead become greater than benchmark set by PLN, i.e. 2%.
5. Conclusion
The results show that electricity consumption in East Java is depend on hours, days, and national holidays. Forecasting the electricity consumption in East Java using time series regression with the addition of lag as predictor produces the smallest forecast error values. Otherwise, ARIMAX Quantile Regression model that involves lag and dummy variables as predictors yield more accurate forecast than the model without lag and dummy variables as predictors. Generally, based on the evaluation criteria, it could be concluded that the ARIMAX Quantile Regression model yields better forecast performance than the time series regression model. Thus, the proposed method can be a potential alternative to forecast short-term electricity in East Java. Moreover, both time series regression and ARIMAX-QR methods with additional lag values as predictors could capture accurately the patterns in the data and produce better predictions than those which not using lag of variables.

Acknowledgements
This research was supported by LPPM ITS under scheme of “Penelitian Pemula 2017”, project No. 695/PKS/ITS/2017. The authors gratefully acknowledge the support from ITS for funding the research and anonymous for their useful suggestions.

References
[1] Koenker R and Bassett Jr. G 1978 *Econometrica* **46** 33-50
[2] Ma L and Pohlman L *The European Journal of Finance* **14** 409-425
[3] Zhang G P 2003 *Neurocomputing* **50** 159 – 175
[4] Taylor J W 2000 *Journal of Forecasting* **19** 299-311
[5] Xu Q, Tsui K-L, Jiang W and Guo H 2016 *Qual. Reliab. Engng. Int.* **32** 2751–2759
[6] Arunraj N S and Ahrens D 2015 *International Journal Production Economics* **170** 321-335
[7] Soares L J and Medeiros M C 2008 *International Journal of Forecasting* **24** 630-644
[8] Wei W W S 2006 *Time Series Analysis Univariate and Multivariate Methods* (New York : Pearson Education, Inc.)
[9] Suhartono, Lee M H and Hamzah N A 2010 *Proceedings of the Regional Conference on Statistical Sciences* 30-41
[10] Draper N R and Smith H 1998 *Applied Regression Analysis, 3rd Edition* (New York: John Wiley & Sons, Inc.)
[11] Koenker R 2005 *Quantile Regression* (New York: Cambridge University Press)
[12] Cade B S and Noon B R 2003 *Frontiers in Ecology and the Environment* **1** 8 412–420
[13] Hyndman R J and Koehler A 2006 *International Journal of Forecasting* **22** 679-688