Hierarchical time series bottom-up approach for forecast the export value in Central Java

D A Mahkya, B S Ulama, and Suhartono
Department of Statistics, Faculty of Mathematics and Natural Sciences
Sepuluh Nopember Institute of Technology, Surabaya, Indonesia
E-mail: danialmahkya@gmail.com, \{brodjol_su, suhartono\}@statistika.its.ac.id

Abstract. The purpose of this study is Getting the best modeling and predicting the export value of Central Java using a Hierarchical Time Series. The export value is one variable injection in the economy of a country, meaning that if the export value of the country increases, the country’s economy will increase even more. Therefore, it is necessary appropriate modeling to predict the export value especially in Central Java. Export Value in Central Java are grouped into 21 commodities with each commodity has a different pattern. One approach that can be used time series is a hierarchical approach. Hierarchical Time Series is used Bottom-up. To Forecast the individual series at all levels using Autoregressive Integrated Moving Average (ARIMA), Radial Basis Function Neural Network (RBFNN), and Hybrid ARIMA-RBFNN. For the selection of the best models used Symmetric Mean Absolute Percentage Error (sMAPE). Results of the analysis showed that for the Export Value of Central Java, Bottom-up approach with Hybrid ARIMA-RBFNN modeling can be used for long-term predictions. As for the short and medium-term predictions, it can be used a bottom-up approach RBFNN modeling. Overall bottom-up approach with RBFNN modeling give the best result.

1. Introduction
Export is a method or activity to market goods produced domestically and sold or marketed abroad. If a country open international trade and become an exporter of goods, the domestic producers of goods that will benefit and domestic consumer goods will be harmed. The value of exports is one variable injection in the economy of a country, meaning that if a country’s exports increases, the country’s economy will increase even more [1]. The economic growth of Central Java province after 1999 have always had a positive growth is also determined by several indicators that tend to increase. Foreign sector plays an important role in the economy of Central Java province. During the period 1997 to 2010 the value of exports and imports increased. [2] Every year in Central Java province exports continued to rise, means the economy in Central Java will continue to grow and will certainly have an impact on Indonesia’s export sector in general. Predicting the export value, especially in Central Java province is certainly much needed for policy development. Because with an approximate value in the future, will help in the future careful planning. Moreover, the value of exports in the province of Central Java, which tends to increase from year to year, so it needs a correct policy that these conditions persist. If policymakers about how the development of exports has predictive value of exports in Central Java province in the future, the policy created will be able to anticipate the phenomena that occur in the pattern of export Central Java province. Therefore, it is necessary an appropriate
predictive models to determine whether export in Central Java province will continue to rise or it will go down. Research on the value of exports of Central Java never been done before, among other things by Rejekiningsih in 2012 studied the concentration of exports in Central Java province. Data of export value in Central Java has a hierarchical structure or data that have levels. The value of exports in total as a data hierarchy level 0 or level up while the value of exports in commodities were 21 commodity as a data hierarchy level 1 or lower levels. In the data analysis used hierarchical level 1 series as much as 19 because there are some data that can not be modeled. Data of export value in Central Java for each commodity has a pattern of stable or unstable. It is therefore we model and forecast all series in hierarchy individually at all level using ARIMA, RBFNN and Hybrid ARIMA-RBFNN which is expected to capture the phenomena that occurs in the data pattern. Selection of the best model in this study using sMAPE.

2. Methods
According to Wei (2006) time series is a sequence of observations. Usually sorted by time, at a certain moment a few sorted by time interval. One of the important things in building a time series model is able to predict the value of a time series for the future [4]. By predicting the value of a time series that will come, may be obtained new stage in an analysis, which is the stage of policy making. Policy in terms of economic, political and social.

2.1. Hierarchical Time Series
The structure of Hierarchical Time Series can be illustrated in Fig. 1 [5]:

![Figure 1. Hierarchical Time Series Structure](image_url)

Based on Fig.1 variable Total is the time series data level 0 which is the sum of time series data at level 1. Likewise for time series data level 1 which is the sum of the time series data at level 2 and so on until level to K. In general notation for Hierarchical Time Series approaches, used Yi,t consisting of all observations to a level i and time t. All observations on all time t is denoted by \( Y_t = [Y_{1,t}, Y_{2,t}, ..., Y_{K,t}] \)' is a column vector [6]. So it can be written as follows: \( Y_t = SY_{K,t} \) the S is the sum of the order m x mK matrix which is the sum of all the lower level time series data. The concept of Buttom-up approach is to generate the prediction result data for each of the time series data at a lower level and then summing to obtain total predictions or predictions of top level [5]. The advantages of the approach Button-up is no information is lost during the process of incorporation for each level. But on the other hand, Button-up approach probably would not be accurate to predict if there is missing data or noisy at lower levels [6].

2.2. Autoregressive Integrated Moving Average (ARIMA)
There are 3 stages that are used to build an ARIMA model such as [7] ARIMA model tentatively identified by analyzing the data of the past, estimating the model parameters are unknown,
checked diagnostic residuals generated to determine the adequacy of the model or indicate whether or not to do repair. Once those steps are completed, then the best ARIMA model is obtained. General model of ARIMA \((p, d, q)\) is as follows [3]:

\[
(1 - \phi_1 B - \ldots - \phi_p B^p)(1 - B)^d Y_t = \theta_0 + (1 - \theta_1 B - \ldots - \theta_q B^q) a_t
\]

### 2.3. Radial Basis Function Neural Network (RBFNN)

RBFNN appear as one type of neural network around the 80s. Construction of RBFNN contains three distinct layers. Input Layer The first layer is derived from the sensor units. The second layer is the Hidden Layer with high dimensional RBF which aims to process many layer Perceptron. The third layer is the output layer of the network provides a response to the activation pattern is applied to the Input Layer [8]. RBFNN have architecture as follows [9]:

![Figure 2. Architecture of RBFNN with input p and m neurons in the hidden layer.](image)

where \(\phi_j(x) = \exp\left(-\frac{||x-\mu||^2}{2\sigma_j^2}\right)\)

Based on Fig. 2 there are \(p\) neurons in the input layer and nonlinear processed with gaussian activation function. Inputs are grouped into several groups so that the average and variance of each group can be determined.

### 2.4. Hybrid Model

Time Series forecasting is a very important aspect of Forecasting in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship [10]. Various models have been developed time series analysis, from the univariate models to multivariate models. Therefore models to the analysis of time series is very popular and is often used for forecasting. Meanwhile Neural Network (NN) is one method that can be used in forecasting. NN is a highly flexible modeling. By NN, no special model in its establishment. The resulting model adapting the patterns generated from the data [10]. Both models are very much used for forecasting, but not between the two models are suited to the circumstances of various phenomena. Therefore [10] proposed a Hybrid model for forecasting by considering the autocorrelation linear structure and nonlinear components with the following equation:

\[
Y_t = L_t + N_t
\]

where \(L_t\) is a component of linear and nonlinear component is \(N_t\). The two components are estimated from the data using each method. For the first model representing linear components, the model will generate residuals from linear models containing nonlinear relationships.
2.5. Accuracy of Prediction

In measuring the accuracy of prediction, often used Mean Absolute Percentage Error (MAPE). But the value of MAPE would be problematic if the data has a value close to zero or equal to zero [11]. Therefore, in this study used Symmetric Mean Absolute Percentage Error to measure the accuracy of the model by the formula:

\[
sMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i - \hat{Y}_i|}{(|Y_i| + |\hat{Y}_i|)/2} \times 100
\]  

(3)

At the stage of checking the accuracy of prediction used the Rolling Forecast Origin approach with illustration shown in Fig 3 and Fig 4.

Here are the stages of Rolling Forecast Origin:
- Determine the best model is used based on the training data.
- Estimate based on the best model and re-estimate based on best model in observation \( n+i \) with \( i = 0,1,2 \ldots, N-n \) where \( n \) is the number of observations in which the training data and \( N \) is the number of overall data.
- Calculate the accuracy of the predictions for each \( i \).

Based on Fig 3 and Fig 4, the blue dot is re-estimated data, while the red dot is the predicted outcome for each \( h \). Where \( h \) is the prediction of the future.

3. Methodology

The data used in this study is secondary data obtained from the Berita Resmi Statistik (BRS) Badan Pusat Statistik province of Central Java with the period from January 2007 until December 2015. In the analysis of this study, there are several steps that must be completed are as following:

(i) ARIMA modeling for each variable in level 0 and level 1 with the following steps:
   (a) Transforming or differencing if not stationary time series data.
   (b) Identify the ARIMA model based on ACF and PACF pattern.
   (c) Estimate the parameter.
   (d) Check the white noise assumption for residuals.
   (e) Getting the best model.
   (f) Calculate forecast value

(ii) RBFNN modeling for each variable in level 0 and level 1 with the following steps:
   (a) Decide the input to the model RBFNN based on ACF and PACF patterns.
   (b) Decide the mean an variance using K-means Cluster.
   (c) Getting the best criteria of some mean and variance.
   (d) Calculate forecast value.
(iii) Hybrid ARIMA-RBFNN modeling for each variable in level 0 and level 1 with the following steps:

(a) ARIMA Modeling for each variable in level 0 and level 1 and get the predicted value.
(b) Produce the Residuals ARIMA of each variable in level 0 and level 1.
(c) RBFNN Modeling for every Residuals ARIMA in level 0 dan level 1.
(d) Calculate forecast value with 
\[
\hat{Y}_t = \hat{L}_t + \hat{N}_t
\]

(iv) Calculating a prediction hierarchy bottom-up approach with ARIMA, RBFNN and Hybrid modeling.

4. Result and Discussion
The result of individual modelling with ARIMA, RBFNN and Hybrid ARIMA-RBFNN show in Table 1, Table 2 and Table 3. On selecting the best model, used Hierarchical Time Series Bottom-up approach with ARIMA, RBFNN and Hybrid ARIMA-RBFNN modeling using Origin Rolling Forecast. The result is shown in Fig 5 and Table 4 is based on criteria of sMAPE.

| Series | ARIMA Order | Series | ARIMA Order |
|--------|-------------|--------|-------------|
| Total  | ARIMA (2,0,0) | 11 | ARIMA (2,1,0) |
| 1      | ARIMA (0,1,1) | 12 | ARIMA (1,0,1)(0,1,1) |
| 2      | ARIMA (1,1,1) | 13 | ARIMA (0,1,1) |
| 3      | ARIMA (0,1,1) | 14 | ARIMA (0,1,1) |
| 4      | ARIMA (0,1,1) | 15 | ARIMA (0,1,1) |
| 5      | ARIMA (0,1,1) | 16 | ARIMA (0,1,1) |
| 6      | ARIMA (0,1,1) | 17 | ARIMA (1[20],0,0) |
| 7      | ARIMA (1[10],1,0) | 18 | ARIMA (0,1,1[23]) |
| 8      | ARIMA (2,0,0) | 19 | ARIMA (0,1,1) |
| 9      | ARIMA (0,1,1) |     |             |
| 10     | ARIMA (1,1,1) |     |             |

Figure 5. Selection the best model for Level 0
Table 2. The Result of RBFNN Modeling

| Series | RBFNN Model | Series | RBFNN Model |
|--------|-------------|--------|-------------|
| Total  | 2 Input 5 Neuron* | 11  | 3 Input 4 Neuron* |
| 1      | 2 Input 6 Neuron* | 12  | 2 Input 3 Neuron* |
| 2      | 2 Input 2 Neuron* | 13  | 2 Input 4 Neuron* |
| 3      | 2 Input 6 Neuron* | 14  | 2 Input 4 Neuron* |
| 4      | 2 Input 4 Neuron* | 15  | 2 Input 6 Neuron* |
| 5      | 2 Input 5 Neuron* | 16  | 2 Input 3 Neuron* |
| 6      | 2 Input 3 Neuron* | 17  | 2 Input 3 Neuron* |
| 7      | 2 Input 6 Neuron* | 18  | 2 Input 2 Neuron* |
| 8      | 2 Input 3 Neuron* | 19  | 2 Input 3 Neuron* |
| 9      | 3 Input 6 Neuron* |
| 10     | 2 Input 6 Neuron* |

* Neuron in Hidden Layer

Table 3. The Result of Hybrid in RBFNN Process

| Series | RBFNN Model | Series | RBFNN Model |
|--------|-------------|--------|-------------|
| Total  | 2 Input 6 Neuron* | 11  | 2 Input 4 Neuron* |
| 1      | 2 Input 6 Neuron* | 12  | 2 Input 2 Neuron* |
| 2      | 2 Input 6 Neuron* | 13  | 2 Input 4 Neuron* |
| 3      | 2 Input 6 Neuron* | 14  | 2 Input 2 Neuron* |
| 4      | 2 Input 3 Neuron* | 15  | 2 Input 5 Neuron* |
| 5      | 2 Input 2 Neuron* | 16  | 2 Input 3 Neuron* |
| 6      | 2 Input 5 Neuron* | 17  | 2 Input 6 Neuron* |
| 7      | 2 Input 4 Neuron* | 18  | 2 Input 4 Neuron* |
| 8      | 2 Input 2 Neuron* | 19  | 2 Input 3 Neuron* |
| 9      | 2 Input 6 Neuron* |
| 10     | 2 Input 3 Neuron* |

* Neuron in Hidden Layer

Based on Fig 5 and Table 4 it can be concluded that the level 0, i.e. Total Export Value, Bottom-up approach with RBFNN modeling can be used for prediction of short and medium term. While the bottom-up approach with hybrid ARIMA-RBFNN modeling can be used for long-term predictions. Overall bottom-up approach with RBFNN modeling give the best result. The next step is to predict the value of exports in Central Java province with the chosen model. The results of the predictive value of total exports Central Java is shown in Fig 6.

Based on Fig 6, it can be explained that the forecasting result coincide with the actual data. But in 12-step ahead forecast shows the result with slight fluctuations.

5. Conclusion
Based on the results and discussion, then obtained some conclusions. First based of Fig 5 and Table 4, for a total export value bottom-up approach with hybrid ARIMA-RBFNN modeling
Table 4. sMAPE value for total export value (level 0)

| h  | Bottom-up | Individual |
|----|-----------|------------|
|    | ARIMA     | RBFNN      | HYBRID     | ARIMA     | RBFNN      | HYBRID     |
| 1  | 6.15      | 7.39       | 6.28       | 6.73      | 9.75       | 6.31       |
| 2  | 6.58      | 6.15       | 6.78       | 7.50      | 10.86      | 7.31       |
| 3  | 7.87      | 6.29       | 8.18       | 8.50      | 13.01      | 8.55       |
| 4  | 8.25      | 6.37       | 8.59       | 9.73      | 13.01      | 10.03      |
| 5  | 7.87      | 5.41       | 8.20       | 9.64      | 12.47      | 9.77       |
| 6  | 8.03      | 6.24       | 8.61       | 9.81      | 10.12      | 9.81       |
| 7  | 6.12      | 5.23       | 6.80       | 7.62      | 8.45       | 7.69       |
| 8  | 2.71      | 3.03       | 3.54       | 4.47      | 10.72      | 4.39       |
| 9  | 2.04      | 3.06       | 1.88       | 2.33      | 10.96      | 1.97       |
| 10 | 4.16      | 4.79       | 4.39       | 4.76      | 11.88      | 4.46       |
| 11 | 5.67      | 6.66       | 5.43       | 5.50      | 13.11      | 5.68       |
| 12 | 1.82      | 0.91       | 0.54       | 1.51      | 17.43      | 2.06       |

Mean 5.61 5.13 5.77 6.51 11.81 6.50

Figure 6. Predicted total export value (level 0) for 12 step ahead using Bottom-up with ARIMA modeling

provides the best results for long-term predictions, while for the short and medium term may use a bottom-up approach with RBFNN modeling. Overall bottom-up approach with RBFNN modeling give the best result. The actual data and forecast values for 12 step ahead show in Fig 6.

References

[1] Nursiah C 2011 Peranan ekspor dalam perekonomian Indonesia Jurnal Ekonomi 19 1
[2] Tri W R 2012 Konsentrasi ekspor provinsi Jawa Tengah Jurnal Ekonomi Kuantitatif Terapan 5 pp 109–118
[3] Wei W W S 2006 Time Series Analysis Univariate and Multivariate Method 2nd Edition (New York: Pearson Education)
[4] Jonathan D C and Kung-Sik C 2008 Time Series Analysis with Applications in R (Lowa City, Lowa: Springer)
[5] George A, Ahmed, Roman A & Rob J H 2007 Hierarchical Forecasting for Australian Domestic Tourism (Monash University)

[6] Lin H S and Peter W F S 2013 Grouped time-series forecasting with an application to regional infant mortality counts Centre for Population Change Working Paper 40

[7] Douglas C M, Cheryl J and Murat K 2008 Introduction to Time Series Analysis and Forecasting (Hoboken, New Jersey: Wiley)

[8] Zhigang Z and Jun W 2010 Advances in Neural Network Research and Applications (Berlin: Springer)

[9] Grady H 2011 Artificial Neural Networks in Biological and Environmental Analysis (New York: CRC Press)

[10] Zhang G P 2003 Time series forecasting using a hybrid ARIMA and neural network model Neurocomputing 50 pp 159–175

[11] Jan G D G and Rob J H 2006 25 years of time series forecasting Int. J. of Forecasting 22 pp 443–473