COMPARISON OF EEMD AND MODIFIED EEMD METHOD IN PREDICTING OF CURLY RED CHILI PRICE IN INDONESIA*

Zilrahmi Zilrahmi[1], Hari Wijayanto[2], Farit M. Afendi[3], Rizal Bakri[4‡]

[1]Department of Statistics, Universitas Negeri Padang (UNP), Indonesia, zilrahmi@fmipa.unp.ac.id
[2]Department of Statistics, IPB University, Indonesia, hari@apps.ipb.ac.id
[3]Department of Statistics, IPB University, Indonesia, fmafendi @apps.ipb.ac.id
[4]Department of Accounting, STIEM Bongaya, Indonesia, rizal.bakri@stiem-bongaya.ac.id
[‡]corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)
Vol 4 No 2 (2020), 374 - 381

Abstract

The fluctuations of curly red chili prices affect the inflation rate in Indonesia. So, the basic characteristics of price movement and correct prediction for curly red chili prices become concerns in various studies. Empirical Mode Decomposition (EMD) method helps to examine behavioral characteristics of curly red chili prices in Indonesia easily. Ensemble EMD (EEMD) and modified EEMD are the decomposition method of time series which is the development of the EMD method. The decomposed data with EMD methods can also be used for price forecasts. The forecasting with ARIMA and trend polynomial performed to assess the effect of decomposition with EMD methods for forecast stability of curly red chili price in Indonesia under various conditions. The results show that the ability of a decomposition method to produce the actual components that describe the pattern of data signals affect the accuracy of the predicted value obtained using the model. The predicted value using the decomposed data by modified EEMD always better than EEMD on the overall condition.

Keywords: arima, EEMD, modified EEMD, trend polynomial.

1. Introduction

Chili's utilization as a food ingredient, for health, and as industrial raw materials make chili become one of the horticultural crops that is always needed by Indonesian people. The demand for chili affects the fluctuation of chili's price. Curly red chili is a commodity

* Received Des 2019; Accepted Jul 2020; Published online on Jul 2020

with the highest level fluctuation than other types. The fluctuations affect the inflation rate in Indonesia so that the basic characteristics of price movement and correct prediction for curly red chili prices become concerns in various studies.

The price patterns are quite complex because the fluctuations are usually irregular and difficult to estimate. Consequently, the analysis became complicated, especially for a long time series data pattern. In statistics, the pattern analysis for the price behavior of chili can be done with data decomposition (Subagyo, 2000). The decomposition of the time series data is to separate the elements that influence the activities of data. In other words, the time series is broken down into several components then each component is identified separately to help understand the behavior of the data series (Makridakis et al., 1999).

Huang et al. (1998) offer Empirical Mode Decomposition (EMD) for extracting the various component in a set of time-series data. EMD can parse a data signal into mutually orthogonal components called Intrinsic Mode Functions (IMFs). Wu & Huang (2009) suggested the ensemble approach to address the issue of mixing modes arising from the use of EMD. Ensemble EMD (EEMD) is done by add noise to the original data and repetition before obtaining final IMFs. The averaging process to obtain a final IMFs on EEMD cause no longer mutually orthogonal of final IMFs. Kuo et al. (2013) proposed the concept of re-analysis clump for recombining the IMFs from EEMD. Then, Wijayanto et al. (2015) take modification to the approach taken by Kuo to obtain components that reflect the actual pattern and independent of each other. Fransiska et al. (2014) had made a good prediction of data decomposition with the EEMD method using ARIMA. In this study, we want to compare the results of predictions against the decomposition data using EEMD and modified EEMD methods.

2. Material and Methods

2.1 Forecasting Methods

a. Ensemble Empirical Mode Decomposition

Ensemble Empirical Mode Decomposition (EEMD) is the development of the Empirical mode decomposition method conducted by Wu & Huang (2009) to deal with the symptoms of mixing mode, which is a condition where an IMF resulting from decomposition contains patterns with large scale differences. The final decomposition component in the EEMD process is obtained by calculating the average of the ensemble performed.

The number of ensemble members can be determined as many as 100 trials (Zhang et al., 2010) with a standard deviation of the white noise between 0.1 or 0.2 (Zhang et al., 2008).

b. Modified EEMD

Wijayanto et al. (2015) modified the approach taken by Kuo et al. (2013) to obtain components that are independent and reflect true patterns. Modifications were made in the process of grouping IMFs produced at an iterative stage in the EEMD. The measure of distance used in the clustering is the correlation between averages.

c. Autoregressive Integrated Moving Average

Forecasting methods are generally used for two purposes, namely to analyze the data series and the selection of forecasting models that best match the data series. The Integrated Moving Average Autoregressive (ARIMA) is a forecasting method introduced by Box-Jenkins in 1970. ARIMA can predict time series data based only on the behavior of observed variable data, so this method is most popular for forecasting univariate time series data. Box-Jenkins method models consist of non-seasonal stationary models: AR(p), MA(q) and ARMA(p,q), non-stationary and non-seasonal models: ARI(p,d), IMA(d, p) dan ARIMA(p, d, q) and seasonal model SARIMA(p, d, q)(P, D, Q)S. ARIMA (p, d, q) model is

$$\phi_p(B)(1-B)^d x(t) = \theta_q(B)a(t)$$

where p is: autoregressive order, d is Integrated, q is moving average order, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\theta_q(B) = 1 - \theta_1 B - \theta_1 B^2 - \cdots - \theta_q B^q$, and $(1-B)^d$ is an order of non-seasonal distinctions.

d. Trend Polynomial

The trend line is the regression line with the independent variable *t* is a time variable. A common form of regression polynomial is as follows:

$$y = a_0 + a_1 t + a_2 t^2 + \cdots + a_N t^N$$

with $a_1, a_2, \dots, a_N$ are coefficients of the polynomial, and *N* is non-negative integers.

2.2 Methods

1) Split the data into training and testing with some variations (Figure 1).

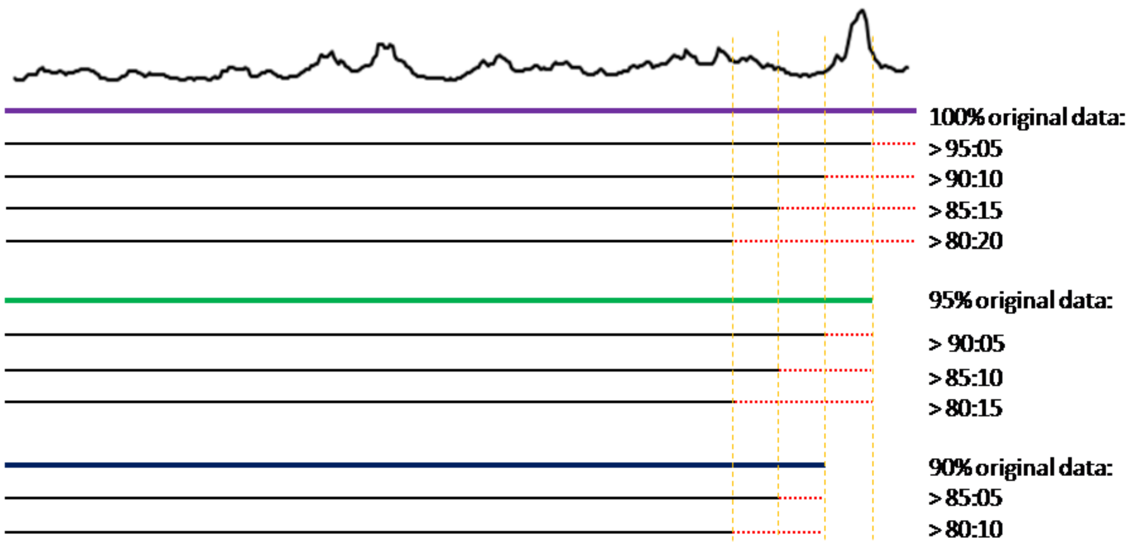

Figure 1: Distribution of training and testing data.

2) Decompose the training data become IMFs and residue using EEMD algorithms;
   - Add a series of white noise to original data to make new data series $x(t)$.
   - Identify all local extreme; maxima and minima of x(t).
   - Create an upper envelope $\mathrm{e_{max}(t)}$ and the lower envelope $\mathrm{e_{min}(t)}$, through all the local maxima and minima which are associated with cubic spline interpolation.
   - Calculate mean, $m_1(t) = \frac{(e_{max}(t)+e_{min}(t))}{2}$
   - Extract detail as IMF candidate, $d_1(t) = x(t) - m_1(t)$
   - Analyze detail as an IMF: Function has the same number of zero-crossings and extreme or different only one, symmetrical function to local zero mean.
   - If the detail is not an IMF then repeat steps (2)-(6), with $d_1(t)$ as *x*(*t*) in the next iteration, $d_1(t) - m_{11}(t) = d_{11}(t)$. Repeat this step until the IMF's criteria are fulfilled. If after *k* iterations a detail is an IMF, $d_{1(k-1)}(t) - m_{1k}(t) = d_{1k}(t)$, then the first IMF obtained by the formula $c_1(t) = d_{1k}(t)$.
   - Extract the residue, $r_1(t) = x(t) - c_1(t)$.
   - Check the residue as a monotonous function (do not have an extreme value). Steps (2)-(8) repeated *i* iterations if the residue is not a monotonous function. If the residue is a monotonous function then the sifting process is stopped, and $r(t) = r_{i-1}(t) - c_i(t)$, $r_0(t) = x(t)$ with 1 < i < M.
   - Steps (1)-(9) repeated *j* times with different white noise in each iteration.
   - Calculate the mean of corresponding IMFs, $c_i(t) = \frac{1}{N}\sum_{j=1}^{N} c_{ij}(t)$, and residue $r = \frac{1}{N}\sum_{j=1}^{N} r_j$ as the final result, *j* is the number of iterations (1, 2, ..., *N*), and *i* is an index of the IMF.

3) Decompose the training data become IMFs and residue using modified EEMD algorithms;
   - Decompose the data using the EEMD algorithm.
   - Grouping the iterations that produce the same amount of IMFs into one group.
   - Obtain final IMFs for each group *p*, *p = (1, 2, ..., P):*
     - $\mathrm{IMF}_{i(p)}$ is the mean of IMFs with same index *i* of all member n_p of group *p*, $IMF_{i(p)} = \frac{1}{n_p}\sum_{j=1}^{n_p} IMF_{ij(p)}$, $i = 1, 2, \ldots, m_p$.
     - Do hierarchical cluster analysis to IMFs, the distance of $\mathrm{IMF}_{i(p)}$ and $\mathrm{IMF}_{h(p)}$ is $1 - abs\left(corr\left(IMF_{i(p)}, IMF_{h(p)}\right)\right)$. Two IMFs with distance less than 0,8 are combined into a single IMF.
   - Combine IMFs between groups with a pair-matching process. Each pair consists of the IMF of a group and the IMF another group. Do it sequentially, starting with Group 1 and Group 2 at the first. The results from this group then combined with Group 3, and so on (Wijayanto et al., 2015).

4) Obtain IMFs results to the training data using EEMD and modified EEMD algorithm.
5) Determine the best model by ARIMA for all IMFs components; Obtain a polynomial trend model that is appropriate to the overall residue.
6) Predicting with best ARIMA model for the IMFs, and polynomial trend towards the residue. The predicted value for decomposition is the sum of all IMFs and residue.
7) Calculate Mean Absolute Percentage Error (MAPE) predicted value of training and testing data to see the goodness of the model used in predicting, $MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x(t)-\hat{x}(t)}{x(t)}\right| \times 100\%$ (Bakri et al., 2019).
8) Perform (2-7) steps for all variation of the data distribution to see the effect of the decomposition method used for forecasting stability.

3. Result and Discussion

The data is divided into training data to obtain models and testing data to evaluate the goodness of the model. The goodness of forecasting heavily depends on testing data series used to evaluate the goodness of the models. The very volatile of weekly curly red chili price data can lead the data pattern used as testing the data do not describe the data pattern used as training data. It can lead the model does not provide good results in forecasting the testing data although the data is derived from the same series. By this issue, to see the stability of data forecasting for decomposed data, a vary partition of training and testing data performed as follows 95:05, 90:10, 85:15, 80:20, 90:5, 85:10, 80:15, 85:5 and 80:10.

Overall the best model in Table 1 is then used to predict each component of decomposition. The prediction results of each IMFs and residue then totaled to obtain the prediction value of the weekly curly red chili price. Then, MAPE calculated as shown in Table 2.

The lower value of MAPE indicates the prediction value getting closer to the actual value. Table 2 shows that the value of MAPE on data decomposition using EEMD modified first is always smaller, it means the decomposition using modified EEMD produced the ARIMA and polynomial regression prediction value that always stable closer to the actual value than EEMD in all variations condition of the data.

The decomposition performed to improve the accuracy of forecasting and understanding the behavior of the data series (Makridakis et al., 1999). The ability of modified EEMD to extract the data into IMFs depicting actual patterns of a signal (Wijayanto et al., 2015), make the forecasting of decomposed weekly curly red chili price getting better. It also can be seen from the graph the prediction value of each decomposed data for each partition in Figure 2. Based on Figure 2 it can be said that the pattern of prediction value obtained from the decomposed data follows the pattern of the actual data. The prediction value of the decomposed data using modified EEMD is closer to the actual values and have a pattern more closely resembling the pattern of the actual data than using EEMD.

Table 1: The best model of each component decomposition.

Number of Training data	decomposition method	Component	Model
95%	EEMD	IMF1	ARIMA (2,0,2)
		IMF2	ARIMA (3,0,2)
		IMF3	ARIMA (3,0,0)
		IMF4	ARIMA (3,0,0)
		IMF5	ARIMA (3,1,0)
		IMF6	ARIMA (3,2,0)
		residue	$y = 7.e^{-6}t^4 - 0.004t^3 - 0.44t^2 + 37.48t + 15772$
	Modified EEMD	IMF1	ARIMA (2,0,1)
		IMF2	ARIMA (4,0,0)
		IMF3	ARIMA (3,2,0)
		residue	$y = 6.e^{-6}t^4 - 0.003t^3 + 0.38t^2 + 35.93t + 16050$
90%	EEMD	IMF1	ARIMA (1,0,1)
		IMF2	ARIMA (5,0,0)
		IMF3	ARIMA (3,0,0)
		IMF4	ARIMA (3,0,0)
		IMF5	ARIMA (3,1,0)
		IMF6	ARIMA (2,1,0)
		residue	$y = -0.13t^2 + 71.04t + 15441$
	Modified EEMD	IMF1	ARIMA (1,0,3)
		IMF2	ARIMA (4,0,0)
		IMF3	ARIMA (3,0,0)
		residue	$y = -0.13t^2 + 71.43t + 15408$
85%	EEMD	IMF1	ARIMA (1,0,3)
		IMF2	ARIMA (5,0,0)
		IMF3	ARIMA (3,0,0)
		IMF4	ARIMA (3,0,0)
		IMF5	ARIMA (3,0,0)
		IMF6	ARIMA (2,2,0)
		residue	$y = -0.07t^2 + 58.14t + 15880$
	Modified EEMD	IMF1	ARIMA (1,0,3)
		IMF2	ARIMA (4,0,0)
		IMF3	ARIMA (3,0,0)
		residue	$y = -0.06t^2 + 55.19t + 16056$
80%	EEMD	IMF1	ARIMA (3,0,4)
		IMF2	ARIMA (5,0,0)
		IMF3	ARIMA (3,0,0)
		IMF4	ARIMA (3,0,0)
		IMF5	ARIMA (4,2,0)
		IMF6	ARIMA (2,2,0)
		residue	$y = -0.06t^2 + 54.05t + 16041$
	Modified EEMD	IMF1	ARIMA (3,0,4)
		IMF2	ARIMA (4,0,0)
		IMF3	ARIMA (3,0,0)
		IMF4	ARIMA (3,1,0)
		residue	$y = 4e^{-6}t^4 - 0.003t^3 + 0.54t^2 + 5.73t + 17121$

Table 2: MAPE for testing prediction value for all data partition.

Scope of Data	**100%**				**95%**			**90%**	
Training: Testing	**95%:5%**	**90%:10%**	**85%:15%**	**80%:10%**	**90%:5%**	**85%:10%**	**80%:15%**	**85%:5%**	**80%:10%**
EEMD	170.7	**33.4**	31.8	39.1	41.9	**37**	41.3	41.9	**43.9**
Modified EEMD	111.4	**26.3**	31.2	37.2	36.7	**36.2**	37.2	40.6	**38.9**

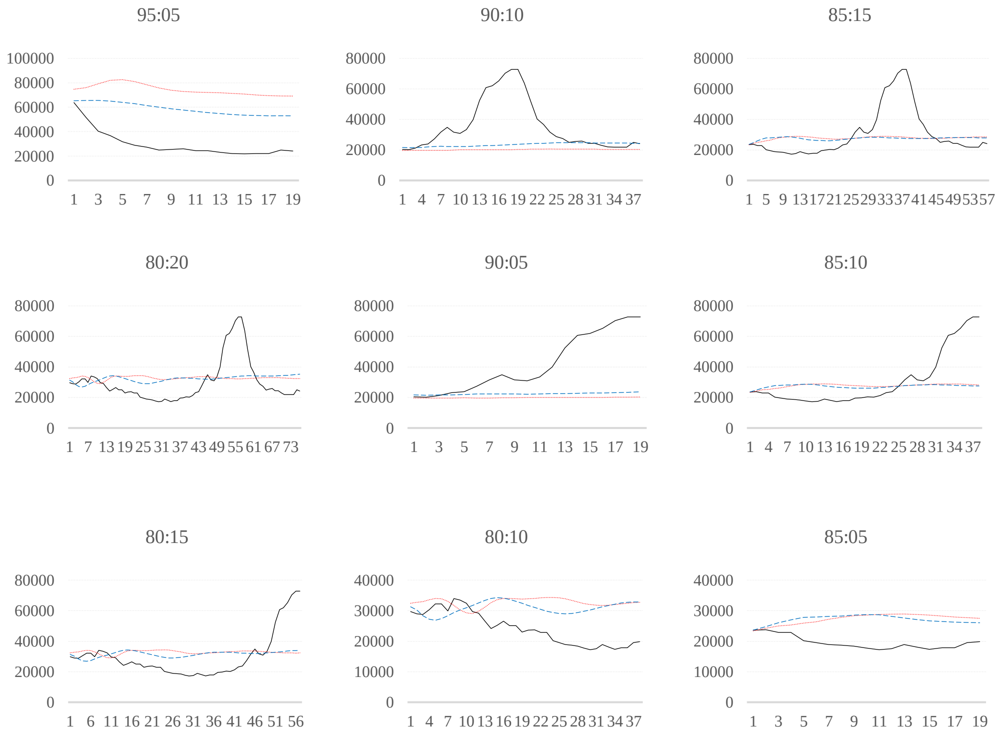

Figure 2: Graph of prediction value using the best models of the various conditions of the data. ——actual data, – – – EEMD, - - - - modified EEMD.

4. Conclusion

The decomposition helps to understand the behavior of the weekly curly red chili price in Indonesia easily. In this case, the ability of a decomposition method to produce the actual components that describe the pattern of data signals affect the accuracy of the predicted value obtained using the model. The accuracy of the model in predicting testing data from decomposed data with modified EEMD always better than EEMD on the overall condition. Data decomposition using the modified EEMD method provides nearly approximates prediction value and a more closely resembling pattern of the actual testing data than using EEMD.

References

Bakri, R., Data, U., & Astuti, N. P. (2019). Aplikasi Auto Sales Forecasting Berbasis Computational Intelligence Website untuk Mengoptimalisasi Manajemen Strategi Pemasaran Produk. *JSINBIS (Jurnal Sistem Informasi Bisnis)*, *9*(2): 244–251.

Fransiska, H., Wijayanto, H., & Sartono, B. (2014). Ensemble Decomposition Method for Predicting the Price of Rice in Jakarta. *IOSR Journal of Mathematics*, *10*(4): 92–97.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., … Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *454*(1971): 903–995.

Kuo, C.-Y., Wei, S.-K., & Tsai, P.-W. (2013). Ensemble empirical mode decomposition with supervised cluster analysis. *Advances in Adaptive Data Analysis*, *5*(01): 1–19.

Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1999). Metode dan aplikasi peramalan. *Jakarta (ID): Erlangga*.

Subagyo, P. (2000). Forecasting: konsep dan aplikasi. *Yogyakarta(ID): BPFE UGM*.

Wijayanto, H., Sartono, B., Fitrianto, A., & Nursyifa, C. (2015). Improving the Components Independence of Decomposition in Time Series Data. *Far East Journal of Mathematical Sciences*, *96*(3): 303–314.

Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, *1*(01): 1–41.

Zhang, J., Yan, R., Gao, R. X., & Feng, Z. (2010). Performance enhancement of ensemble empirical mode decomposition. *Mechanical Systems and Signal Processing*, *24*(7): 2104–2123.

Zhang, X., Lai, K. K., & Wang, S.-Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, *30*(3): 905–918.