Ensemble Decomposition Method for Predicting the Price of Rice in Jakarta

Herlin Fransiska[1], Hari Wijayanto[1], Bagus Sartono[1]
[1]Department of Statistics, Bogor Agricultural University, Indonesia

Abstract:*Time series analysis is one of statistical procedures in time series data which is applied to predict the conditions that will come in the context of decision making. Generally, the huge size of data not only non linear but also non stationary and it is difficult to be interpreted in concrete. This problem can be solved by performing the decomposition process, the process of changing into a simpler form. Decomposition method that is Ensemble Empirical Mode Decomposition (EEMD). Decomposed time series data can also be used for prediction of the initial data. The ensemble methods can be used such as Fourier analysis used because IMF patterned sinusoid and ARIMA is used because this method is very popular in time series data. The methodology is applied to forecast weekly rice prices in Jakarta province from January 2002 to August 2013.*
Keyword: *ARIMA, EEMD, Ensemble, Fourier Analysis, Time Series data.*

I. Introduction

Time series analysis is one of statistical procedures in time series data which is applied to predict the conditions that will come in the context of decision making. Some methods of time series data analysis are ARIMA, VAR, ARCH, GARCH. Generally, the huge size of data not linear and also not stationary and it is difficult to be interpreted in concrete. This problem can be solved by performing the decomposition process, the process of changing into a simpler form. Decomposition method that is currently growing rapidly in time series data is Empirical Mode Decomposition (EMD) by Huang et al in 1998 [1]. EMD is empirical and analytical techniques are adaptive to the processing of data, particularly data not linear and also not stationary [2]. However, EMD is not able to settle the emergence of mixing mode, so the presence of Ensemble Empirical Mode Decomposition method (EEMD) is proposed to address this weakness [3]. Ensemble is a method in improving the ability of a method. The accuracy of prediction is an important part in the analysis of time series data. A predictive model which has a high predictive power can be obtained with ensemble technique [4].

Decomposed time series data can also be used for prediction of the initial data. Prediction is carried out in every IMF and residual component, then all the results of predictions are summed to obtain the initial prediction of time series data. In an effort to get the accurate components' prediction, hybrid ensemble technique can be used where it is using a variety of modeling methods and then combine the predictions generated by each method into a final prediction. Several methods can be used such as Fourier analysis and ARIMA. The selections of these models are based on either Fourier analysis used because IMF patterned sinusoid and ARIMA is used because this method is very popular in time series data.

This study applies ensemble decomposition technique to the data of daily rice prices in Jakarta province from January 2002 to August 2013 which is converted into weekly data. This study uses EEMD method to decompose the data, then used a hybrid ensemble technique to predict the price of rice. Fourier analysis and ARIMA is the ensemble method here.

II. Methodology

2.1. Data

The data used in this study were collected from Perum Bulog and the Ministry of Trade Directorate Staples and Strategic Goods. The data was about the price of rice in January 2002-August 2013 (the data length of 606 weeks). The data then divided into two parts, the data in January 2002 until the 2nd week of June 2013 (594 weeks) for the modeling and the data in 3rd week of June 2013 to August 2013 (12 weeks) for validation. Software used in this study was Ms. Excel 2007, Minitab16 and R.3.0.1.

a. Methods

EMD algorithm
1. Identifying all local extreme value, the maximum and minimum of the input data time series, x(t) with $1 < t < n$. A local maximum value at period-t can be known if $x(t) > x(t-1)$ and $x(t) > x(t+1)$. In contrast, the value of the local minimum period-t is determined if $x(t) < x(t-1)$ and $x(t) < x(t+1)$.

2. Making the top edge, $e_{max}(t)$ and the bottom edge, $e_{min}(t)$ through the local extreme points of maximum and minimum cubic spline interpolation associated with.
3. Calculating the average $m_1(t)=(e_{min}(t)+e_{max}(t))/2$.
4. Extracting detail as will the IMF, $d_1(t)=x(t)-m_1(t)$.
5. Analyzing the fulfillment of a condition of the IMF detail as follows:
   i.   The function has a number of zero-crossings and extreme same lot or just a different one. A zero-crossing can be identified if the $d_1(t)<0<d_1(t-1)$ or $d_1(t)>0>d_1(t-1)$.
   ii. The function is symmetric with respect to the local zero mean (local zero mean).
6. If the detail is not an IMF then repeat steps a-e, subsequent sifting, by setting d(t) as x(t) new in the next iteration, $d_1(t)-m_{11}(t)=d_{11}(t)$. This process continues until the formation of an IMF meeting the criteria. If the detail is an IMF after k iterations, $d_{1(k-1)}(t)-m_{1k}(t)=d_1(t)$,, then the IMF1 obtained through the formula$c_1(t)=d_{ik}(t)$.
7. Extract residue, $r_1(t)=x(t)-c_1(t)$.
8. Checking residue as a monotonic function (do not have extreme values). If the residue is not a monotonic function then repeat steps 1-7 (sifting) as much as i iterations. If residues including monotonic function of the sifting process is stopped $r(t)=r_{i-1}(t)-c_i(t)$ with $r_0(t)=x(t)$ and $1<i<M$ [1].

EEMD algorithm
1. The addition of a series of white noise in the data.
2. Decompose the data that had been given white noise with EMD algorithm.
3. Iteratively repeating steps 1-2 as N, but with a different white noise in each iteration.
4. Calculating the Fund $c_i(t)$ and residual r (defined as a trend) of the ensemble mean with the formula $c_i(t)=\frac{1}{N}\sum_{j=1}^{N}c_{ij}(t)$ and $r=\frac{1}{N}\sum_{j=1}^{N}r_j$[3] [5] [6].

ARIMA algorithm
1. Create a plot of data and a formal test to identify stationary series.
2. If the data has been stationary, make ACF and PACF graphs to determine the initial ARIMA model.
3. Do the parameter estimates ARIMA (p, d, q) according to the initial model defined in section b. Then test the feasibility of the model.
4. Making prediction some periods ahead using the best models [7] [8] [9].

Modified Fourier Analysis Algorithm
1. Compiling the data of each IMF based on its period (p), with t=t+p, for t= ±1, ±2, ….
2. Calculating the average of IMF in each period (p).
3. Using the average for predicting IMF with $\hat{z}_{t+p}=\hat{z}_t$.
4. Making prediction some periods ahead.

Methods
1. Making plots of weekly rice-price data.
2. Extracting a number of IMF and residue based on Ensemble EMD algorithm.
3. Determining the contribution of each IMF which was produced to the input data by using the averaging period, Pearson correlation and variance ratio percentage.
4. Making prediction IMFs and residu using arima methods.
5. Making prediction IMFs and residu using modified fourier analysis methods.
6. Making prediction weekly rice-price using ensemble methods (mean of modified fourier analysis and ARIMA methods).

III.   Results And Discussion

3. 1. Data Exploration

Generally, the weekly prices of rice in Jakarta from January 2002 to August 2013 have a positive trends pattern, it can be seen from Figure 1, but sometimes the prices tend to be stable just like in 2002 and 2005. This rice price is neither linear nor stationary not only in its mean but also in the variance so that decomposition with EEMD method was taken. EEMD represented the signal of nonstationary and nonlinear with an assumption that signal was composed from different simple local oscillation.
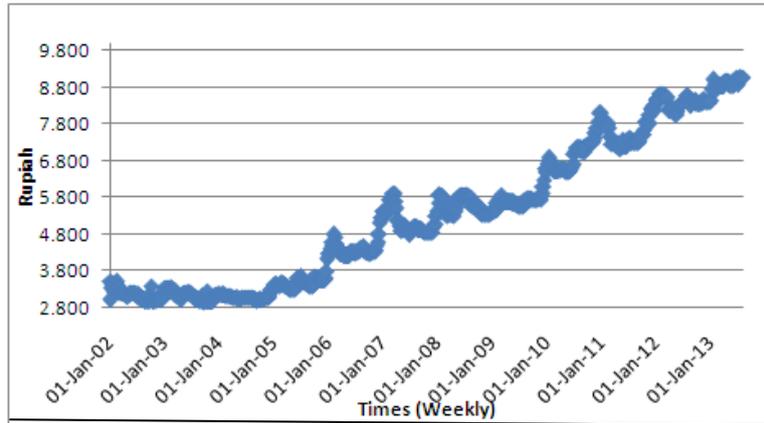
Figure 1 the weekly prices of rice in Jakarta from January 2002 to August 2013

Decomposition of rice price data by using EEMD decomposed a signal into 6 IMF and residue, IMF was extracted from high frequency to low frequency as shown in Figure 2.
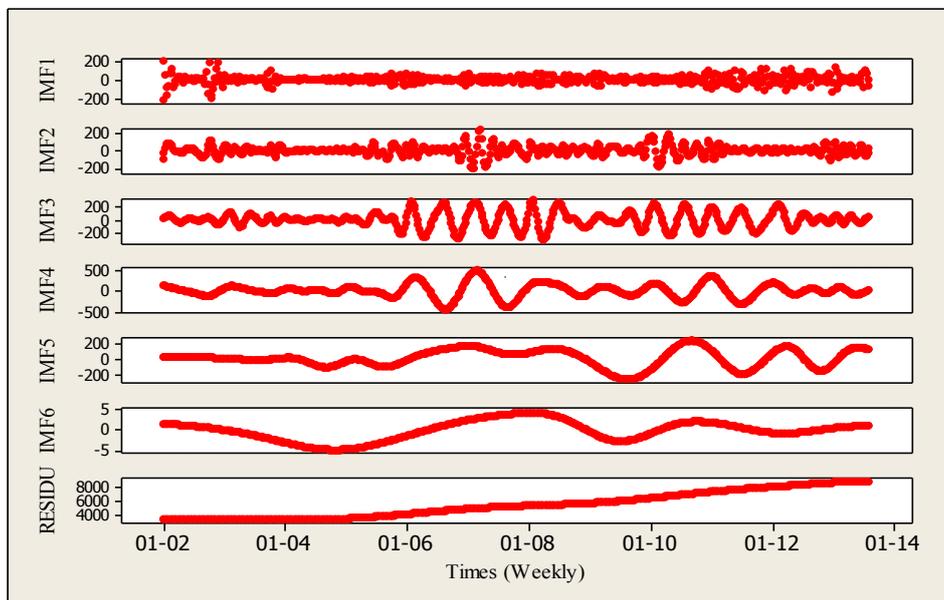
Figure 2 Results of rice price decomposition by using EEMD

Table 1 presents the number of peaks, the number of valleys, averaging period, correlation and variance of each IMF. The number of valleys and peaks which was sequentially smaller made the average period sequentially getting bigger. Correlations tend to be greater. The correlation reflected the relevance of IMF and residual component of the weekly price data of rice in Jakarta. Residual component had a very large correlation, 0.99. This suggests that the movement direction of the residual component mostly in accordance with the direction of rice price. Based on the ratio of variance in Table 1, IMF1 and IMF2 gave the smallest contribution, IMF4 contributed 0.80% which was the largest contribution of IMF components, while the residual component gave the largest contribution (97.70%) of all components.

Table 1 Description of IMFs and Residu

	Number Of Peaks	The Number Of Valleys	Averaging Period	Correlation Pearson	Variance	Variance Ratio
Data					3816393.32	
IMF1	167	166	4	0.04	1982.31	0.10
IMF2	73	73	8	0.03	3419.02	0.10
IMF3	32	32	19	0.06	15233.16	0.40
IMF4	14	15	43	0.08	28897.53	0.80
IMF5	8	7	76	0.13	13990.90	0.40
IMF6	3	3	202	0.36	5.99	0.00
Residu				0.99	3727969.70	97.00
Total						99.50

3. 2. Prediction with ARIMA

IMF components and residues were predicted by using ARIMA. Based on the data plot, ACF and PACF plots, and a formal test that dickey-fuller test and Bartlett test, showed that IMF components were stationary in its mean and the variance.

Component	Models	AIC	SBC
IMF1	ARIMA(1,0,3)	10.21645	10.24603
IMF2	ARIMA(2,0,3)	7.41233	7.44936
IMF3	ARIMA(2,0,2)(0,0,1)[13]	3.91074	3.94777
IMF4	ARIMA(2,0,2)(1,0,0)[14]	6.29400	6.33102
IMF5	ARIMA(2,0,2)	-1.65500	-1.62538
IMF6	ARIMA (2,0,2)	-12.2392	-12.2096
Residu	ARIMA (2,0,2)	-4.11270	-4.08308

Table 2 The best models of each component

The best models in Table 2 were used to predict the IMF component 12 weeks ahead. The prediction results could be seen in Table 3. The Prediction in Table 3 ensemble in order to get a prediction of Jakarta weekly rice prices for 12 weeks ahead. The ensemble concept used was adding it up because it had been known that the sum of IMF components and the residue will reproduce the original data without any distortion or missing information. This was also applied in the prediction results where the total amount prediction of IMF components and the residue will generate predictions of rice price data. Prediction of rice price could be seen in Table 4.

Times	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	Residu
21-May-13	9.47102E+00	-44.854	-18.882	-97.282	156.051	0.744	8846.493
28-May-13	1.07755E+00	-8.704	-34.805	-96.378	157.251	0.753	8853.255
04-Jun-13	-3.60289E+00	21.552	-46.098	-93.713	157.524	0.760	8859.976
11-Jun-13	6.33699E-01	34.634	-51.036	-89.194	156.959	0.766	8866.656
18-Jun-13	-6.71032E-01	30.306	-50.415	-81.961	155.650	0.769	8873.293
25-Jun-13	-2.69217E-01	15.265	-45.713	-72.370	153.700	0.770	8879.888
02-Jul-13	-3.92963E-01	-1.486	-38.111	-61.007	151.210	0.770	8886.441
09-Jul-13	-3.54853E-01	-12.849	-28.857	-48.030	148.290	0.767	8892.952
16-Jul-13	-3.66590E-01	-15.888	-19.370	-34.414	145.048	0.763	8899.419
23-Jul-13	-3.62975E-01	-11.789	-10.856	-21.336	141.594	0.757	8905.844
30-Jul-13	-3.64089E-01	-4.215	-3.266	-9.643	138.035	0.748	8912.226
06-Aug-13	-3.63746E-01	2.851	3.747	0.494	134.479	0.738	8918.564

Table 3 Predicted component of the IMF and residue in 12 Weeks

Times	$\hat{x}_t$	x_t	$\left\lvert\dfrac{x_t - \hat{x}_t}{x_t}\right\rvert$
21-May-13	8851.7408	8840.0000	0.0013
28-May-13	8872.4496	8840.0000	0.0037
04-Jun-13	8896.3974	8840.0000	0.0064
11-Jun-13	8919.4185	8916.1900	0.0004
18-Jun-13	8926.9722	8923.8100	0.0004
25-Jun-13	8931.2707	9054.2860	0.0136
02-Jul-13	8937.4252	8863.8100	0.0083
09-Jul-13	8951.9187	9030.4760	0.0087
16-Jul-13	8975.1920	9048.5710	0.0081
23-Jul-13	9003.8513	9060.0000	0.0062
30-Jul-13	9033.5218	9060.0000	0.0029
06-Aug-13	9060.5087	9060.0000	0.0001
	MAPE		0.0050

Table 4 Prediction of Jakarta weekly rice prices with ARIMA

Based on Table 4, the evaluation of prediction results by using MAPE showed that the result prediction using ARIMA had been very good because MAPE obtained at 0.005%.

3.3 Prediction with Fourier Analysis

The second method used a modification of Fourier analysis. IMF component was predicted by using the average at each period. IMF1 to IM6 had periods (p) 4, 8, 19, 43, 76, 202. Then, each component of IMF will apply $\hat{Z}_t = \hat{Z}_{t+p}$ obtained from the average in each period. On average in each IMF period was used to predict the IMF component in 12 weeks ahead. The prediction results could be seen in Table 5. Next, the ensemble performed in order to get a prediction of weekly rice prices in Jakarta 12 weeks ahead as the ARIMA. Final prediction by this method could be seen in Table 6.

Times	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
21-May-13	-3.610	-7.582	-9.432	4.867	1.947	-3.098
28-May-13	-6.884	-2.693	-9.376	13.016	7.347	-3.057
04-Jun-13	2.582	3.921	-8.410	20.871	12.724	-3.013
11-Jun-13	5.561	6.985	-6.460	28.265	18.051	-2.964
18-Jun-13	-3.610	5.105	-3.726	35.001	23.300	-2.912
25-Jun-13	-6.884	0.892	-0.750	40.874	28.449	-2.856
02-Jul-13	2.582	-4.998	2.142	45.674	33.474	-2.797
09-Jul-13	5.561	-8.741	4.598	49.255	38.353	-2.734
16-Jul-13	-3.610	-7.582	6.316	56.677	43.067	-2.668
23-Jul-13	-6.884	-2.693	7.179	57.352	47.594	-2.599
30-Jul-13	2.582	3.921	7.068	56.980	51.916	-2.527
06-Aug-13	5.561	6.985	6.513	55.644	56.012	-2.451

Table 5 Predicted component of the IMF in 12 Weeks with Fourier Analysis

Times	$\hat{x}_t$	x_t	$\left\|\dfrac{x_t - \hat{x}_t}{x_t}\right\|$
21-May-13	8829.510	8840.000	0.001186675
28-May-13	8851.269	8840.000	0.001274757
04-Jun-13	8887.819	8840.000	0.005409444
11-Jun-13	8914.532	8916.190	0.000185963
18-Jun-13	8923.923	8923.810	0.000012667
25-Jun-13	8935.879	9054.286	0.013077490
02-Jul-13	8957.328	8863.810	0.010550574
09-Jul-13	8972.352	9030.476	0.006436469
16-Jul-13	8982.773	9048.571	0.007271674
23-Jul-13	8994.738	9060.000	0.007203336
30-Jul-13	9018.646	9060.000	0.004564438
06-Aug-13	9030.581	9060.000	0.003247139
MAPE			0.005035052

Table 6 Prediction of Jakarta weekly rice prices with Fourier Analysis

Based on Table 6, the evaluation of prediction results by using MAPE for this method showed that the results had been very good prediction when MAPE obtained at 0.0050.

3.4 Ensemble Prediction

Final Prediction of weekly rice price in Jakarta on observations 595-606 (12 Weeks) was obtained by ensemble final prediction of ARIMA and Fourier analysis as shown in Table 7. Based on Table 7, the evaluation of the prediction results by using MAPE showed very good prediction because MAPE result was 0.0049.

Times	$\hat{x}_t$	x_t	$\left\|\dfrac{x_t - \hat{x}_t}{x_t}\right\|$
21-May-13	8840.6254	8,840.00	0.00007
28-May-13	8861.8593	8,840.00	0.00247
04-Jun-13	8892.1082	8,840.00	0.00589
11-Jun-13	8916.9753	8,916.19	0.00009
18-Jun-13	8925.4476	8,923.81	0.00018
25-Jun-13	8933.5749	9,054.29	0.01333
02-Jul-13	8947.3766	8,863.81	0.00943
09-Jul-13	8962.1354	9,030.48	0.00757
16-Jul-13	8978.9825	9,048.57	0.00769
23-Jul-13	8999.2947	9,060.00	0.00670
30-Jul-13	9026.0839	9,060.00	0.00374
06-Aug-13	9045.5449	9,060.00	0.00160
MAPE			0.00490

Table 7 Final Prediction of Jakarta weekly rice prices
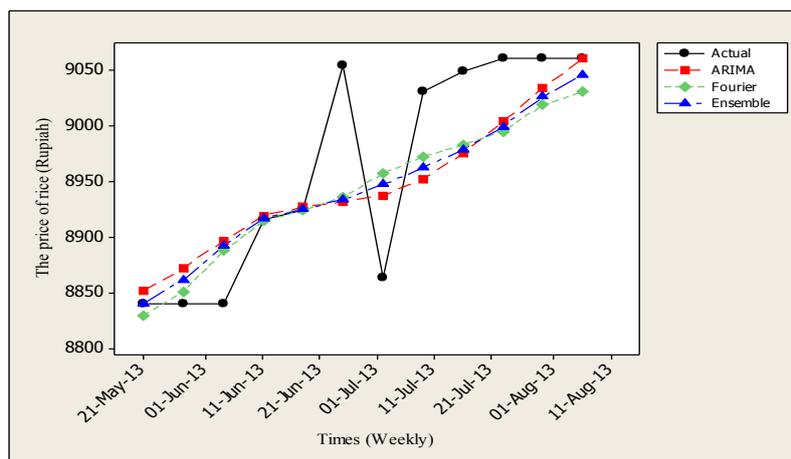
Figure 3 Plot Data Jakarta Weekly Rice Price and Prediction Results

Based on Figure 3, it can be seen that the plots of rice price data and the results of the forecast 12 weeks ahead of ARIMA method, Fourier analysis and ensemble. Figure 10 provides information that the overall ensemble decomposition method will give a very good prediction.

IV.   Conclusion

Data of rice prices that has been decomposed tend to be easier to interpret and produce excellent predictions. The goodness of prediction can be seen from MAPE which is relatively small on a weekly rice price data in Jakarta. In the rice price data in Jakarta from January 2002 to August 2013, the prediction by using ARIMA and Fourier analysis produces good predictions those are 0.005 and 0.005. Ensemble hybrid in this prediction produces MAPE 0.0049. In general, the price of rice weekly prediction methods in Jakarta by using hybrid ensemble techniques through EEMD process produces predictions with a very good accuracy.

References

[1].   Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc. Roy. Soc. Lond., A(454):903–995.
[2].   Nursyifa C. 2013. Identifikasi Pola Pergerakan Harga Beras Melalui Dekomposisi Deret Waktu Secara Ensemble. Bogor : IPB.
[3].   Huang NE, Wu Z. 2005. Ensemble empirical mode decomposition: A noise assisted data analysis method. Advances in Adaptive Data Analysis. 1(1):1-41.
[4].   Zhu M. 2008. Kernels and ensembles: Perspectives on statistical learning. The American Statistican 62 (2): 97-109.
[5].   Wu Z, Huang NE. 2004. A study of the characteristics of white noise using the empirical mode decomposition method. Proc. Roy. Soc. Lond., A(460):1597-1611.
[6].   Zhang J, Yan R, Gao RX Feng Z. 2010. Performance Enhancement Of Ensemble Empirical Mode Decomposition. Mechanical System And Signal Processing. 24: 2104-2123.
[7].   Box GEP, Jenkins GM. 1976.Time Series Analysis: Forecasting and Control, Holden-day Inc., San Francisco, CA.
[8].   Cryer JD, Chan K. 2008. Time Series Analysis With Applications in R. Springer. New York.
[9].   Wei WWS. 2006. Time Series Analysis Univariate and Multivariate Methods. Second Edition. Pearson Education, Inc. US.