

Prediction of Poor Inhabitant Number Using Least Square and Moving Average Method

Ningrum Ekawati

School of Information Management and Computer “AMIKOM Yogyakarta”, STMIK AMIKOM Yogyakarta
Yogyakarta, Indonesia

Abstract—The number of poor inhabitant in South Kalimantan decreased within the last three years compared with the previous years. The numbers of poor inhabitant differs from time to time. This scaled dynamical number has been a problem for the local government to take proper polices to solve this matter. It will then be necessary to predict a potential number of poor inhabitants in the next year as the basis on subsequent policy making. This research will apply both Least Square and Moving Average method as the measurement to count prediction values. From the results of the study, the prediction analysis by using those two methods is valid for predicting acquired number of poor inhabitant for the next period according to the data from the previous year. Based on the study, the validity of Least Square was 98.35% and Moving Average was 98.79% by using the data in the last seven years.

Keywords—Poor Inhabitant; Prediction; Least Square; Moving Average

I. INTRODUCTION

Poverty is the main problem in South Kalimantan Province, Indonesia [1]. According to Statistic Center Agency, the poor inhabitant is defined as those who averagely spend below the poverty line per capita per month [2]. Based on the Statistic Center Agency in South Kalimantan [2], the number of poor inhabitant decreased each year. In 1999 the number of poor inhabitant was 440,200 and while at the end of 2014 became 182,876 inhabitants. From those data, the number of poor inhabitant decreased 5.28% on average each year and in the last three years decreased 3.31%. There are a decreasing numbers of poor inhabitants in South Kalimantan due to the numbers of building.

The problem was in predicting the number of poor inhabitant in South Kalimantan that decreased in the last three years. However, the prediction cannot be predicted for the next years, and it caused the local government had difficulty to make decision. This study was aimed to decide the way of predicting the number of poor inhabitant in South Kalimantan for the years to come by using Least Square and Moving Average method. Hopefully, the result of the research could help the government to increase the people’s life quality.

II. RESEARCH METHOD

Time series analysis was a statistical analysis method applied to predict a future condition. To make an accurate data, the prediction was conducted for a very long time and much data was needed. As one of the choices to describe a future trend, time series analysis can be applied to reflect dynamic

variable from one time to another [3]. From the previous studies by using Least Square [4-8] and Moving Average [3][9-14] method, the data and analysis showed the future prediction. It was defined as a management process in making decision. It was described as a prediction process in the unknown future situation. In general term, it was well known as a prediction referring to time series estimation or longitudinal type of data [9].

The Least Square method was often used to predict (Y), due to its detail measurement [4]. The trend line (1) was:

$$Y = a_0 + bX \quad (1)$$
$$a = (\Sigma Y) / n$$
$$b = (\Sigma XY) / \Sigma X^2$$

Where:

- Y : Scaled data (time series) = Trend value prediction.
- a₀ : Trend value in the basis year.
- b : Average growing trend value in each year.
- X : Time variable (year).

To conduct the calculation, a certain value in time variable (X) was required so that the total variable score was zero or $\Sigma X = 0$. In analyzing the data with Least Square method, it is generally divided into two parts i.e. “even data” and “odd data” [4].

For odd “n”, where:

The interval between two times was one-unit value

It was marked as negative when it was above 0

It was marked as positive when it was below 0

For even “n”, where:

The interval between two times gains two-unit value

It was marked as negative when it was above 0

It was marked as positive when it was below 0

Generally, linier line equation from time series analysis (2) was:

$$Y = a + bX \quad (2)$$

Description:

Y is a variable that trend was searched.

X is a time variable (year).

Meanwhile, to find constant value “a” and parameter value “b” (3) was:

$$a = \Sigma Y / n, \text{ and } b = \Sigma XY / \Sigma X^2 \quad (3)$$

Moving Average method was a prediction approach by taking some observed groups of the values, finding the average, and using the average values as a prediction of subsequent period. The formula (4) was [10]:

$$F_t = \frac{A_{t-1} + A_{t-2} + A_{t-3} + A_{t-n}}{n} \quad (4)$$

Description:

F_t : Forecast for the coming period
 n : Number of period to be averaged “n”

$A_{t-1} + A_{t-2} + A_{t-3}$: Actual Occurrences in the past period, two period ago, three period ago, and so on respectively.

III. RESULTS AND ANALYSIS

Based on the Statistic Central Agency, the number of poor inhabitant in South Kalimantan from 1996 and not for every year due to the calculation until the Province level for once in three year. So, the data in 1996, 1999, and 2000 was counted every year [2]. So, the prediction calculation of poor inhabitant in this study focused on time year variable. Time series analysis with Least Square and Moving Average can be applied to identify time year variable.

The required data of the research were those of poor inhabitant in South Kalimantan Province starting from 1999 to 2014 [2].

TABLE I. THE DATA OF POOR INHABITANT FROM 1999 TO 2014

No	Year	Total
1	1999	440,200
2	2000	385,300
3	2001	357,500
4	2002	259,800
5	2003	259,000
6	2004	231,000
7	2005	235,700
8	2006	278,451
9	2007	233,500
10	2008	218,898
11	2009	175,977
12	2010	181,963
13	2011	194,623
14	2012	190,597
15	2013	184,297
16	2014	182,876

A. Least Square Method

In this study, the data for the “odd data”, previous collected data from the last nine years were required. Meanwhile, when processing data tabulation for an “even data”, the previous data collection from the last ten years are required.

1) “Odd Data”: Before measuring the prediction of poor inhabitant in 2015, the test was conducted to the number of poor inhabitant in 2008, 2009, 2010, 2011, 2012, 2013, and 2014 to know whether Least Square was valid or not. Compared with the data of poor inhabitant in the last seven years. To find the data of poor inhabitant in 2008, the data of poor inhabitant in 1999 to 2007 was used. The next was finding the value of X, XY, and X².

TABLE II. VARIABLE DATA OF POOR INHABITANT FROM 1999 TO 2007

No	Year	Total (Y)	X	XY	X ²
1	1999	440,200	-4	-1,760,800	16
2	2000	385,300	-3	-1,155,900	9
3	2001	357,500	-2	-715,000	4
4	2002	259,800	-1	-259,800	1
5	2003	259,000	0	0	0
6	2004	231,000	1	231,000	1
7	2005	235,700	2	471,400	4
8	2006	278,451	3	835,353	9
9	2007	233,500	4	934,000	16
Total		2,680,411		-1,419,747	60

Thus, to find “a” value was:

$$a = \Sigma Y / n$$

$$a = 297,823.44$$

And to measure “b” value was:

$$b = \Sigma XY / \Sigma X^2$$

$$b = -23,662.45$$

After gaining the “a” and “b” value, the equation line was:

$$Y = a + bX \text{ (in 2008 the X score was 5)}$$

After measuring the linear line, the number of poor inhabitant in 2008 was:

$$Y = 179,511.192$$

It means that the total number of poor inhabitant in 2008 was 179,511 inhabitants.

The next phase was finding the total number of poor inhabitant in 2009. The data of poor inhabitant in 2000 to 2008 was collected. With the same calculation, in 2009 the result was 187,937 inhabitants. The data of poor inhabitant in 2001 to 2009 was used to identify the number of poor inhabitant for 2010. With the same calculation, the number of poor inhabitant in 2010 was 178,954 inhabitants. The data in 2002 to 2010 was used to know the number of poor inhabitant in 2011. Also, with the same calculation, the total number of poor inhabitant in 2011 was 181,580 inhabitants. The data in 2003 to 2011 was used to identify the number of poor inhabitant in 2012. With the same calculation, the number of poor inhabitant in 2012 was 174,609 inhabitants. The data in 2004 to 2012 was applied to find the number of poor inhabitant in 2013. The same calculation showed that the number of poor inhabitant in 2013 was 171,023 inhabitants. The data in 2005 to 2013 was applied to identify the number of poor inhabitant in 2014. With the same calculation, the total number of poor inhabitant in 2014 was 161,790 inhabitants.

If the different score between Least Square method was >40%, it was considered to be invalid. Compared to the accurate score in 2008, the different was 17.99% (39,387 inhabitants) and the data was valid. In 2009, the different was 6.80% (11,960 inhabitants), it means that the data was valid as well. In 2010, the different was 1.65% (3,009 inhabitants), it means that the data was valid. In 2011 the difference was 6.70% (13,043 inhabitants), it is also means that the data was valid. In 2012, the comparison was 8.39% (15,988 inhabitants), the data also was valid. In 2013, the difference was 7.20%

(13,274 inhabitants), the data was considered to be valid. And in 2014, the comparison was 11.53% (21,086 inhabitants) it means that the data was valid as well. Based on the seven differences, the all data was valid. So, Least Square method was effective or accurate.

The next phase was calculating the prediction number of poor inhabitant in 2015. Based on the data tabulation for “odd data”, the poor inhabitant data were needed from the last nine years, starting from 2006 to 2014.

TABLE III. VARIABLE DATA OF POOR INHABITANT FROM 2006 TO 2014

No	Year	Total (Y)	X	XY	X ²
1	2006	278,451	-4	-1,113,804	16
2	2007	233,500	-3	-700,500	9
3	2008	218,898	-2	-437,796	4
4	2009	175,977	-1	-175,977	1
5	2010	181,963	0	0	0
6	2011	194,623	1	194,623	1
7	2012	190,597	2	381,194	4
8	2013	184,297	3	552,891	9
9	2014	182,876	4	731,504	16
Total		1,841,182		-567,865	60

The table 3, showed that the “a” and “b” values were obtained. To count “a” and “b” values, the following formula was applied:

To find out “a” value was:

$$a = \Sigma Y / n$$

$$a = 204,575.8$$

And to find “b” value was:

$$b = \Sigma XY / \Sigma X^2$$

$$b = -9,464.42$$

After “a” and “b” values were obtained, the linear line was found as follows:

$$Y = a + bX \text{ (for year 2015 the value of X is 5)}$$

After finding the linear measurement, the number of poor inhabitant in 2015 was as follows:

$$Y = 157,253.7$$

It means that the number of poor inhabitant in 2015 was 157,254 inhabitants.

So, the trend analyzing graphic with Least Square method for the different result and the prediction result in the last seven years were:

TABLE IV. THE DATA OF POOR INHABITANT BY LEAST SQUARE METHOD

Year	Actual	Prediction
2008	218,898	179,511
2009	175,977	187,937
2010	181,963	178,954
2011	194,623	181,580
2012	190,597	174,609
2013	184,297	171,023

2014	182,876	161,790
2015		157,254

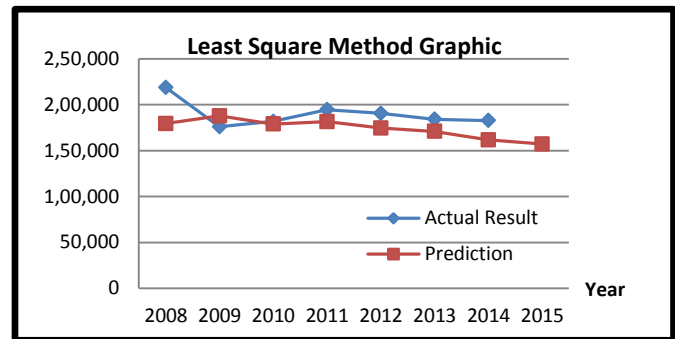


Fig. 1. Graphic Prediction by Using Least Square Method

2) “Even Data”: the required data in the case of “even data” are those of poor inhabitant in South Kalimantan starting from 2005 to 2014.

TABLE V. VARIABLE DATA OF POOR INHABITANT FROM 2005 TO 2014

No	Year	Total (Y)	X	XY	X ²
1	2005	235,700	-9	-2,121,300	81
2	2006	278,451	-7	-1,949,157	49
3	2007	233,500	-5	-1,167,500	25
4	2008	218,898	-3	-656,694	9
5	2009	175,977	-1	-175,977	1
6	2010	181,963	1	181,963	1
7	2011	194,623	3	583,869	9
8	2012	190,597	5	952,985	25
9	2013	184,297	7	1,290,079	49
10	2014	182,876	9	1,645,884	81
Total		2,076,882		-1,415,848	330

Based on table 5, “a” and “b” values were obtained. To find those scores, the following formula was applied:

To find the “a” value was:

$$a = \Sigma Y / n$$

$$a = 207,688.2$$

To find the “b” value was:

$$b = \Sigma XY / \Sigma X^2$$

$$b = -4,290.45$$

After the values of “a” and “b” was gained, the linear measurement was as follows:

$$Y = a + bX \text{ (in 2015 the value of X was 11)}$$

With that equation, the number of poor inhabitant in 2015 was as follows:

$$Y = 160,493.3$$

It means that the prediction number of poor inhabitant was 160,493 inhabitants.

From the calculation of Least Square method, the prediction number of poor inhabitant in 2015 for “odd data” was 157,254 inhabitants. And for the “even data” was 160,493 inhabitants. So, the different was 2.02%. However, the result of

prediction number of poor inhabitant could be wrong due to some cases such as natural disaster, disease epidemic, etc.

B. Moving Average Method

Before measuring the prediction number of poor inhabitant in 2015, the test of number of poor inhabitant in 2008, 2009, 2010, 2011, 2012, 2013, and 2014 by using Single Moving Average method were conducted to know whether the data was valid or not compared with the accurate data of poor inhabitant in the last seven years. To count the prediction number in 2015, the prediction number in 2014 was counted at first.

TABLE VI. THE DATA OF POOR INHABITANT FROM 1999 TO 2014

No	Year	Total
1	1999	440,200
2	2000	385,300
3	2001	357,500
4	2002	259,800
5	2003	259,000
6	2004	231,000
7	2005	235,700
8	2006	278,451
9	2007	233,500
10	2008	218,898
11	2009	175,977
12	2010	181,963
13	2011	194,623
14	2012	190,597
15	2013	184,297
16	2014	182,876

The next phase was defining the value of poor inhabitant by using Single Moving Average method for two periods as follows:

TABLE VII. VARIABLE DATA OF POOR INHABITANT BY SINGLE MOVING AVERAGE TWO PERIOD

No	Year	Actual	Prediction
1	1999	440,200	-
2	2000	385,300	-
3	2001	357,500	412,750
4	2002	259,800	371,400
5	2003	258,960	308,650
6	2004	231,000	259,400
7	2005	235,700	245,000
8	2006	278,451	233,350
9	2007	233,500	257,076
10	2008	218,898	255,976
11	2009	175,977	226,199
12	2010	181,963	197,438
13	2011	194,623	178,970
14	2012	190,597	188,293
15	2013	184,297	192,610
16	2014	182,876	187,447
17	2015	-	183,587

If the different between prediction calculation with Single Moving Average for two periods with the results was >40%, then it is considered to be invalid.

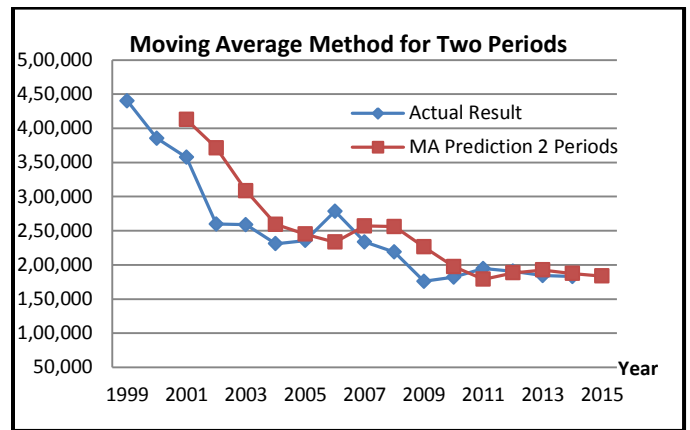


Fig. 2. Graphic Prediction by Using Moving Average Method for Two Periods

Compared to the accurate data in 2008, the different was 14.48% (37,078 inhabitants), it was considered to be valid. In 2009, the different was 22.20% (50,222 inhabitants), it means that the result was considered to be valid. In 2010, the different was 7.84% (15,475 inhabitants), it means that the data was valid. In 2011, the comparison was 8.04% (15,653 inhabitants), the data was valid. In 2012, the difference was 1.21% (2,304 inhabitants), the data was also valid. While in 2013, the difference was 4.32% (8,313 inhabitants) and the data was valid. Lastly, in 2014 the difference was 2.44% (4,571 inhabitants) it also means the data was valid. According to those seven comparisons, the use of Single Moving Average was effective.

C. The Comparison Result of Least Square and Moving Average Method

According to Least Square and Moving Average method, if the difference between prediction calculation with the result was >40% it means that the data were invalid. Based on table 8, there was a comparison result between Least Square and Moving Average method for the last seven years.

TABLE VIII. THE COMPARISON DATA OF LEAST SQUARE METHOD AND SINGLE MOVING AVERAGE METHOD FOR TWO PERIODS

Year	Actual	Prediction		Difference (%)		Result
		Least Square	Moving Average	Least Square	Moving Average	
2008	218,898	179,511	255,976	17.99	14.48	Valid
2009	175,977	187,937	226,199	6.80	22.20	Valid
2010	181,963	178,954	197,438	1.65	7.84	Valid
2011	194,623	181,580	178,970	6.70	8.04	Valid
2012	190,597	174,609	188,293	8.39	1.21	Valid
2013	184,297	171,023	192,610	7.20	4.32	Valid
2014	182,876	161,790	187,447	11.53	2.44	Valid
2015	-	157,254	183,587			

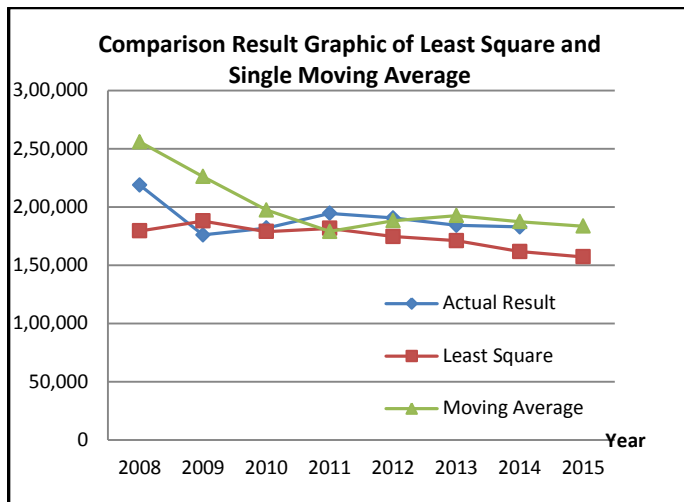


Fig. 3. Comparison Result Graphic of Least Square and Single Moving Average Method for Two Periods

In accordance with the figure 3, the lower data for Least Square method was 1.65% and 1.21% for Moving Average method, so the data was considered to be effective. The validity of Least Square and Moving Average was based on the accurate measured data.

IV. CONCLUSION

The use of Least Square and Single Moving Average method was effective to predict the number of poor inhabitant in South Kalimantan for the next period.

From the result of the prediction calculation, the number of poor inhabitant in 2008 was 179,511 inhabitants. And in 2009 was 187,937 inhabitants. While in 2010 was 178,954 inhabitants. In 2011 was 181,580 inhabitants, in 2012 was 174,609 inhabitants, in 2013 was 171,023 inhabitants, and in 2014 was 161,790 inhabitants. If the difference between the calculation of Least Square and the accurate result was $>40\%$, it was considered to be invalid. Compared to the actual result in 2008, the difference was 17.99% (39,387 inhabitants), it means that the result was valid. In 2009, the difference was 6.80% (11,960 inhabitants), it means that the result was valid. In 2010, the comparison was 1.65% (3,009 inhabitants), it also considered to be valid. In 2011, the comparison was 6.70% (13,043 inhabitants) and the data was valid. In 2012, the difference was 8.39% (15,988 inhabitants) and the data was valid. In 2013, the comparison was 7.20% (13,274 inhabitants) it means that the data was valid. And in 2014, the difference was 11.53% (21,086 inhabitants) and the data was considered to be valid. It means that Least Square method was approximately effective.

From the calculation result from prediction of poor inhabitant in 2008 was 255,976 inhabitants, in 2009 was 226,199 inhabitants, in 2010 was 197,438 inhabitants, in 2011 was 178,970 inhabitants, in 2012 was 188,293 inhabitants, in 2013 was 192,610 inhabitants, in 2014 was 187,447 inhabitants. It was considered to be invalid when the difference between the calculation with Single Moving Average for two periods and the accurate result was $>40\%$. Compared to the accurate data in 2008 the difference was 14.48% (37,078

inhabitants), it means that the result was valid. In 2009, the difference was 22.20% (50,222 inhabitants), and the data was valid. In 2010, the difference was 7.84% (15,475 inhabitants), it means that the result was also valid. In 2011, the comparison was 8.04% (15,653 inhabitants) it means that the data was valid. In 2012 the difference was 1.21% (2,304 inhabitants), and the data was considered to be valid. In 2013, the comparison was 4.32% (8,313 inhabitants), it also considered to be accurate. In 2014, the difference was 2.44% (4,571 inhabitants) it means that the data was valid. Based on the seven comparisons, the all data was accurate or valid. Thus, Single Moving Average was approximately effective.

The accurate result of Least Square was 98.35% and 98.79% for Moving Average, so it was considered to be valid in predicting the number of poor inhabitants.

For the next researches, the number of data and additional variable are required. Smart system can be used as a method to predict the number of poor inhabitant.

REFERENCES

- [1] F. Amina and M. I. Irawan, "Prediksi Jumlah Penduduk Miskin di Kalimantan Selatan Menggunakan Jaringan Syaraf Tiruan Backpropagation," 2014.
- [2] Statistic Center Agency of South Kalimantan Province: <http://kalsel.bps.go.id/Subjek/view/id/23#subjekViewTab1|accordion-daftar-subjek1>
- [3] I. Majerová and T. Pražák, "Estimation of Economic Development in Papua New Guinea: Linear Trend Analysis or Moving Average Model?," *Procedia - Soc. Behav. Sci.*, vol. 110, pp. 450–460, 2014.
- [4] M. I. F. Rambe, "Perancangan Aplikasi Peramalan Persediaan Obat-Obatan Menggunakan Metode Least Square (Studi Kasus: Apotik Mutiara Hati)," pp. 49–53, 2014.
- [5] H. Meilin and X. Yanxia, "Estimation of the complex frequency of a harmonic signal based on a linear least squares method," *Geod. Geodyn.*, vol. 6, no. 3, pp. 220–225, 2015.
- [6] S. A. Korkmaz and M. Poyraz, "Least Square Support Vector Machine and Minimum Redundancy Maximum Relevance for Diagnosis of Breast Cancer from Breast Microscopic Images," *Procedia - Soc. Behav. Sci.*, vol. 174, pp. 4026–4031, 2015.
- [7] S. Wang and W. Shang, "Forecasting direction of China security index 300 movement with least squares support vector machine," *Procedia Comput. Sci.*, vol. 31, pp. 869–874, 2014.
- [8] G. Tan, J. Yan, C. Gao, and S. Yang, "Prediction of water quality time series data based on least squares support vector machine," *Procedia Eng.*, vol. 31, pp. 1194–1199, 2012.
- [9] L. Abdullah, "ARIMA Model for Gold Bullion Coin Selling Prices Forecasting," vol. 1, no. 4, 2012
- [10] R. Kumar and D. Mahto, "Application of Proper Forecasting Technique in Juice Production: A Case Study," vol. 13, no. 4, 2013.
- [11] V. Ruiz, M. A. Pérez, and A. Olasolo, "Dynamic Portfolio Management Strategies based on the Use of Moving Averages," *Procedia - Soc. Behav. Sci.*, vol. 109, pp. 1277–1281, 2014.
- [12] K. Mivule and C. Turner, "Applying Moving Average Filtering for Non-interactive Differential Privacy Settings," *Procedia Comput. Sci.*, vol. 36, pp. 409–415, 2014.
- [13] H. K. Yu, N. Y. Kim, S. S. Kim, C. Chu, and M. K. Kee, "Forecasting the Number of Human Immunodeficiency Virus Infections in the Korean Population Using the Autoregressive Integrated Moving Average Model," *Osong Public Heal. Res. Perspect.*, vol. 4, no. 6, pp. 358–362, 2013.
- [14] A. Pal, J. P. Singh, and P. Dutta, "The Path Length Prediction of MANET Using Moving Average Model," *Procedia Technol.*, vol. 10, pp. 882–889, 2013.