Forecasting Crude Oil Price Using Neural Networks

Komsan Suriya*

Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand

**Corresponding author. E-mail: skomsan@econ.cmu.ac.th*

ABSTRACT

This research constructed the Artificial Neural Networks (Multilayer Feed Forward) to forecast the crude oil price (Brent). The input information was the daily price range between December 27, 2002 to March 18, 2005. Total number of inputs were 561 days. Arranging the input information into groups with 10 consecutive informations in each group, 551 groups were prepared. The model consisted of 10 neurons in the input layer and 1 neuron in the output layer. Quadratic interpolation was utilized to search for the most appropriate number of neurons in the hidden layer. The research question was how many neurons in the hidden layer that would yield the most-accurate forecasting result. The comparisons of models were justified by the 1 day ex ante forecasting results. The Mean Absolute Percentage Error (MAPE) was a measurement of the accuracy. Thirty-four rounds of the forecasting contest were conducted. The least MAPE derived from the best model was 1.98 percent with 200 neurons in the hidden layer.

Key words: Price forecasting, Crude oil, Neural networks

INTRODUCTION

Thailand has risked her energy sector heavily on oil import. The right time to purchase crude oil price will save millions of dollars for Thailand and oil-importing countries. A right notice of an increase of oil price would helpfully suggest the related agents to buy earlier, and vice versa. Thus, the accuracy of the forecasting is so beneficial to agencies who deal with the import of crude oil.

Neural Networks is an efficient tool in training a computer to learn from massive data (Haykin, 1994). It was inspired by the cleverness of the brain cells in the prediction of situations that human faced (Gurney, 1999). It penetrated to economic interest in the time-series analysis (Patterson, 1996; Smith and Gupta, 2000). The technique might promise the satisfied accuracy of the forecasting. Therefore, this study tried to figure out the efficiency of the technique in the forecasting of the crude oil price.

This research aimed to construct an appropriate neural networks model for the forecasting of crude oil price and to test the accuracy of the forecasting. The scope of the study was limited by these following matters. First, only the so-called model "Multilayer Feed Forward" was utilized. Second, only the daily Brent crude oil price was collected. Third, daily forecasting for 34 days was conducted. The last, Mean Absolute Percentage Error (MAPE) was used to measure the accuracy of the forecasting.

METHODOLOGY

A neural networks model consists of at least 3 layers, input layer, hidden layer and output layer. In this study, there are 10 neurons in the input layer, receiving 10 days-lagged price. The daily forecasting forced the output layer to be assigned only one neuron. The hidden layer will be varied to find the best model. The architecture of the model is shown in Fig. 1.

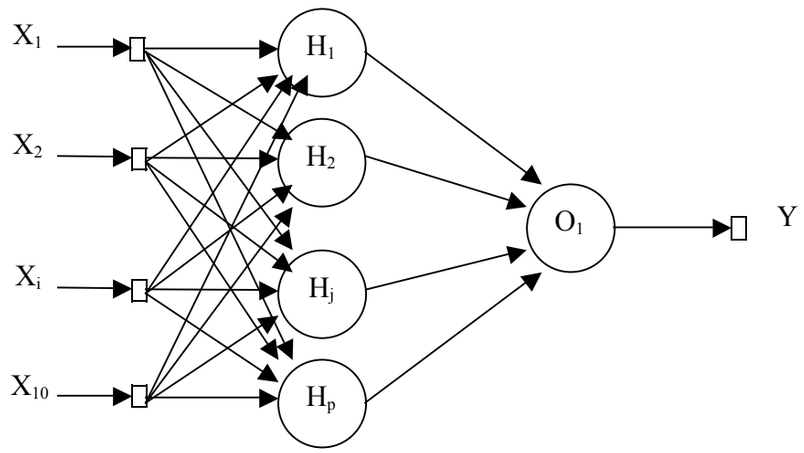

Figure 1. Neural Networks Model for Forecasting Crude Oil Price.

To find out the best model, quadratic interpolation was utilized to search for the minimized error. Mean Absolute Percentage Error (MAPE) was the index to measure the accuracy of each model. The best model will be judged by the least MAPE. The MAPE formula is shown below.

$$MAPE = \frac{|t - y|}{y} \times 100$$

MAPE = Mean Absolute Percentage Error, t = actual price, y = fitted price.

RESULTS AND DISCUSSION

In the first round, three models were constructed. There were 100, 200 and 300 neurons in the hidden layer for each model respectively. These models would produce essential results for the process of quadratic interpolation.

In the second and third round, two more models were suggested by the quadratic interpolation. There were 157 and 231 neurons in the hidden layer for each model respectively. Totally, there were five models for the study. The results of the study are shown in Tables 1-6.

Table 1. The result of forecasting, using the neural networks model with 100 neurons in the hidden layer.

No.	Date	Fitted	Actual	Error (%)	No.	Date	Fitted	Actual	Error (%)
1	21/3/2005	55.64	56.07	0.77	18	14/4/2005	49.37	49.51	0.28
2	22/3/2005	55.33	55.54	0.38	19	15/4/2005	49.69	48.75	1.93
3	23/3/2005	54.46	51.67	5.40	20	18/4/2005	48.77	47.78	2.07
4	24/3/2005	50.39	51.8	2.72	21	19/4/2005	46.41	50.04	7.25
5	28/3/2005	51.20	51	0.39	22	20/4/2005	50.37	50.73	0.71
6	29/3/2005	50.13	50.6	0.93	23	21/4/2005	51.34	50.83	1.00
7	30/3/2005	49.08	49.81	1.47	24	22/4/2005	51.43	52.61	2.24
8	31/3/2005	49.34	52.4	5.84	25	25/4/2005	53.26	52.15	2.13
9	1/4/2005	51.68	53.32	3.08	26	26/4/2005	51.67	51.54	0.25
10	4/4/2005	53.48	55.27	3.24	27	27/4/2005	51.53	50.37	2.30
11	5/4/2005	54.71	53.91	1.48	28	28/4/2005	50.60	49.89	1.42
12	6/4/2005	53.60	53.46	0.26	29	29/4/2005	48.98	49.71	1.47
13	7/4/2005	52.35	51.95	0.77	30	2/5/2005	49.91	49.99	0.16
14	8/4/2005	51.09	51.1	0.02	31	3/5/2005	50.19	48.58	3.31
15	11/4/2005	49.87	50.6	1.44	32	4/5/2005	48.83	49.76	1.87
16	12/4/2005	50.10	50.98	1.73	33	5/5/2005	50.03	49.21	1.67
17	13/4/2005	51.63	49.07	5.22	34	6/5/2005	49.26	48.93	0.67
Summary of the error (%)					**N**	**Max**	**Min**	**Average**	**S.D.**
					34	**7.25**	**0.02**	**1.94**	**1.75**

Source: Calculation

Table 2. The result of forecasting, using the neural networks model with 157 neurons in the hidden layer.

No.	Date	Fitted	Actual	Error (%)	No.	Date	Fitted	Actual	Error (%)
1	21/3/2005	55.997	56.07	0.13	18	14/4/2005	49.425	49.51	0.17
2	22/3/2005	55.67	55.54	0.23	19	15/4/2005	49.975	48.75	2.51
3	23/3/2005	55.618	51.67	7.64	20	18/4/2005	48.814	47.78	2.16
4	24/3/2005	51.131	51.8	1.29	21	19/4/2005	46.68	50.04	6.71
5	28/3/2005	51.81	51	1.59	22	20/4/2005	50.119	50.73	1.20
6	29/3/2005	49.863	50.6	1.46	23	21/4/2005	51.191	50.83	0.71
7	30/3/2005	49.533	49.81	0.56	24	22/4/2005	51.295	52.61	2.50
8	31/3/2005	49.36	52.4	5.80	25	25/4/2005	52.972	52.15	1.58
9	1/4/2005	50.908	53.32	4.52	26	26/4/2005	51.816	51.54	0.54
10	4/4/2005	52.271	55.27	5.43	27	27/4/2005	51.469	50.37	2.18
11	5/4/2005	55.019	53.91	2.06	28	28/4/2005	50.416	49.89	1.05
12	6/4/2005	53.417	53.46	0.08	29	29/4/2005	49.349	49.71	0.73
13	7/4/2005	51.745	51.95	0.39	30	2/5/2005	49.922	49.99	0.14
14	8/4/2005	50.997	51.1	0.20	31	3/5/2005	50.129	48.58	3.19
15	11/4/2005	49.912	50.6	1.36	32	4/5/2005	49.011	49.76	1.51
16	12/4/2005	50.776	50.98	0.40	33	5/5/2005	49.918	49.21	1.44
17	13/4/2005	51.449	49.07	4.85	34	6/5/2005	49.356	48.93	0.87
Summary of the error (%)					**N**	**Max**	**Min**	**Average**	**S.D.**
					34	**7.64**	**0.08**	**1.98**	**2.02**

Source: Calculation

The more neurons in the hidden layer did not always produce the less error. Adversely, it made the model more sensitive to the data. The model adjusted itself too fast. It was good for the monotonic increasing or monotonic decreasing of the price. However, it could not cope well with the pattern of swinging up and down.

This point encouraged a further study to make the model less sensitive when facing the swinging data. An artificial brake should be equipped somehow to the model.

Table 3. The result of forecasting, using the neural networks model with 200 neurons in the hidden layer.

No.	Date	Fitted	Actual	Error (%)	No.	Date	Fitted	Actual	Error (%)
1	21/3/2005	56.03	56.07	0.07	18	14/4/2005	49.73	49.51	0.44
2	22/3/2005	55.96	55.54	0.76	19	15/4/2005	49.61	48.75	1.76
3	23/3/2005	55.41	51.67	7.24	20	18/4/2005	49.23	47.78	3.03
4	24/3/2005	50.19	51.8	3.11	21	19/4/2005	46.83	50.04	6.41
5	28/3/2005	51.74	51	1.45	22	20/4/2005	50.26	50.73	0.93
6	29/3/2005	50.86	50.6	0.51	23	21/4/2005	50.62	50.83	0.41
7	30/3/2005	50.17	49.81	0.72	24	22/4/2005	51.10	52.61	2.87
8	31/3/2005	49.80	52.4	4.96	25	25/4/2005	52.74	52.15	1.13
9	1/4/2005	51.23	53.32	3.92	26	26/4/2005	51.56	51.54	0.04
10	4/4/2005	52.76	55.27	4.54	27	27/4/2005	51.58	50.37	2.40
11	5/4/2005	55.06	53.91	2.13	28	28/4/2005	50.30	49.89	0.82
12	6/4/2005	52.98	53.46	0.90	29	29/4/2005	49.47	49.71	0.48
13	7/4/2005	52.01	51.95	0.12	30	2/5/2005	49.61	49.99	0.76
14	8/4/2005	51.11	51.1	0.02	31	3/5/2005	49.84	48.58	2.59
15	11/4/2005	50.18	50.6	0.83	32	4/5/2005	48.80	49.76	1.93
16	12/4/2005	50.88	50.98	0.20	33	5/5/2005	49.79	49.21	1.18
17	13/4/2005	51.35	49.07	4.65	34	6/5/2005	49.35	48.93	0.86
Summary of the error (%)					**N**	**Max**	**Min**	**Average**	**S.D.**
					34	**7.24**	**0.02**	**1.89**	**1.88**

Source: Calculation

This model was the best model among the five models in the contest. However, this study varied only the number of neurons in the hidden layer. They were not varied in other dimensions such as the number of neurons in the input layer. Without the variation in other dimension, the sense of the best model was limited. A better model might be obtained by such variation which was another encouragement to conduct more experiments.

Table 4. The result of forecasting, using the neural networks model with 231 neurons in the hidden layer.

No.	Date	Fitted	Actual	Error (%)	No.	Date	Fitted	Actual	Error (%)
1	21/3/2005	55.95	56.07	0.26	18	14/4/2005	49.87	49.51	0.37
2	22/3/2005	55.91	55.54	0.54	19	15/4/2005	49.66	48.75	1.91
3	23/3/2005	55.84	51.67	7.81	20	18/4/2005	49.32	47.78	2.72
4	24/3/2005	50.42	51.8	3.88	21	19/4/2005	46.60	50.04	5.93
5	28/3/2005	52.63	51	0.77	22	20/4/2005	50.28	50.73	0.81
6	29/3/2005	50.28	50.6	0.82	23	21/4/2005	50.39	50.83	0.68
7	30/3/2005	49.79	49.81	0.72	24	22/4/2005	51.33	52.61	2.73
8	31/3/2005	49.40	52.4	5.83	25	25/4/2005	52.73	52.15	1.63
9	1/4/2005	51.01	53.32	4.95	26	26/4/2005	51.16	51.54	0.44
10	4/4/2005	52.55	55.27	4.94	27	27/4/2005	51.43	50.37	2.47
11	5/4/2005	54.88	53.91	2.46	28	28/4/2005	50.21	49.89	0.95
12	6/4/2005	51.84	53.46	1.48	29	29/4/2005	49.75	49.71	0.04
13	7/4/2005	52.73	51.95	0.34	30	2/5/2005	49.75	49.99	0.63
14	8/4/2005	50.99	51.1	0.17	31	3/5/2005	49.85	48.58	2.80
15	11/4/2005	49.49	50.6	1.17	32	4/5/2005	48.83	49.76	2.10
16	12/4/2005	51.19	50.98	0.31	33	5/5/2005	49.60	49.21	0.91
17	13/4/2005	51.01	49.07	4.59	34	6/5/2005	49.37	48.93	0.50
Summary of the error (%)					**N**	**Max**	**Min**	**Average**	**S.D.**
					34	**7.81**	**0.04**	**2.02**	**2.00**

Source: Calculation

The quadratic interpolation was utilized twice in this study. The technique required three points to settle a new point. After having the new point, one point from the three original points was eliminated. The remaining three points were used to find out the next point in the second round. This process came to the final round by an insignificant difference between the new point and the original points.

Table 5. The result of forecasting using, the neural networks model with 500 neurons in the hidden layer.

No.	Date	Fitted	Actual	Error (%)	No.	Date	Fitted	Actual	Error (%)
1	21/3/2005	55.95	56.07	0.21	18	14/4/2005	49.87	49.51	2.20
2	22/3/2005	55.91	55.54	0.66	19	15/4/2005	49.66	48.75	0.41
3	23/3/2005	55.84	51.67	8.08	20	18/4/2005	49.32	47.78	3.96
4	24/3/2005	50.42	51.8	2.66	21	19/4/2005	46.60	50.04	0.72
5	28/3/2005	52.63	51	3.19	22	20/4/2005	50.28	50.73	1.87
6	29/3/2005	50.28	50.6	0.64	23	21/4/2005	50.39	50.83	3.21
7	30/3/2005	49.79	49.81	0.04	24	22/4/2005	51.33	52.61	6.87
8	31/3/2005	49.40	52.4	5.72	25	25/4/2005	52.73	52.15	0.89
9	1/4/2005	51.01	53.32	4.33	26	26/4/2005	51.16	51.54	0.87
10	4/4/2005	52.55	55.27	4.92	27	27/4/2005	51.43	50.37	2.43
11	5/4/2005	54.88	53.91	1.80	28	28/4/2005	50.21	49.89	1.11
12	6/4/2005	51.84	53.46	3.04	29	29/4/2005	49.75	49.71	0.74
13	7/4/2005	52.73	51.95	1.50	30	2/5/2005	49.75	49.99	2.09
14	8/4/2005	50.99	51.1	0.22	31	3/5/2005	49.85	48.58	0.64
15	11/4/2005	49.49	50.6	0.21	32	4/5/2005	48.83	49.76	0.08
16	12/4/2005	51.19	50.98	0.66	33	5/5/2005	49.60	49.21	0.48
17	13/4/2005	51.01	49.07	8.08	34	6/5/2005	49.37	48.93	2.62
Summary of the error (%)					**N**	**Max**	**Min**	**Average**	**S.D.**
					34	**8.08**	**0.04**	**2.11**	**1.99**

Source: Calculation

This model was the outer model of the family. In the first round, three original models had to be assigned. It was too much to assign 1,000 neurons in the hidden layer. The computer would process too slow to obtain the forecasting result without any guarantee of the more accuracy. Therefore, 500 neurons seemed to be reasonable for the trading off between the time consumption and the acceptable accuracy. The best model would be often found within this scope of the search.

Table 6. Comparison of Mean Absolute Percentage Error (MAPE) of 5 models.

Model	Number of neurons in the hidden layer	MAPE			
		Max	Min	Average	S.D.
1	100	7.25	0.02	1.94	1.75
2	157	7.64	0.08	1.98	2.02
3	200	7.24	0.02	1.89	1.88
4	231	7.81	0.04	2.02	2.00
5	500	8.08	0.04	2.11	1.99

Source: Calculation

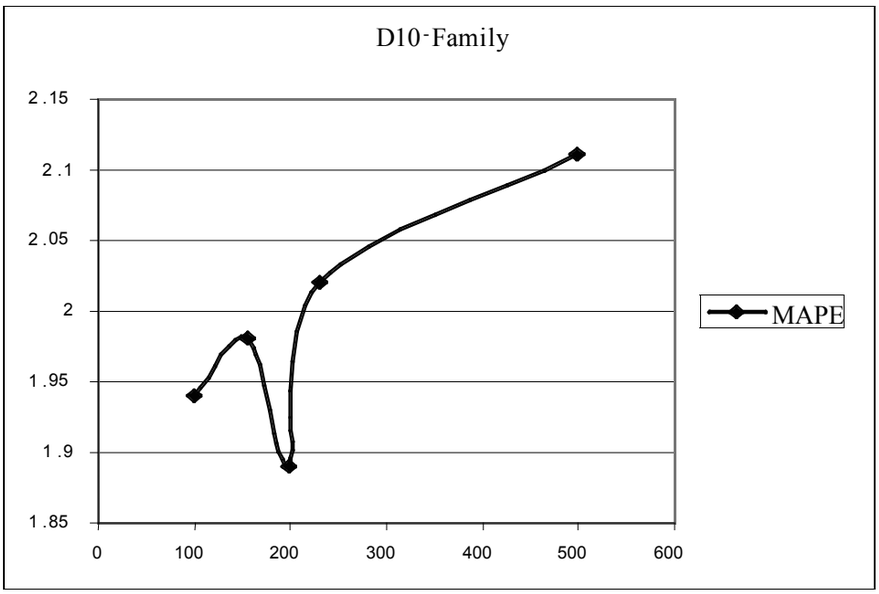

Figure. 2 Summary of the Mean Absolute Percentage Error (MAPE) of 5 models.

The more neurons in the hidden layer would not guarantee the more accuracy. By the way, the best model was mysteriously found among the increasing curve of the MAPE. This was not only the result appeared in this study but also the same with other studies using Neural Networks in forecasting time-series. Such the strange phenomena should be focused and paid more attention to give a better explanation in further studies.

CONCLUSION

The best model for the forecasting of crude oil price occupied 200 neurons in the hidden layer. The model yielded the least error, only 1.89% on average measured by the Mean Absolute Percentage Error (MAPE). It will be interesting to conduct further studies to compare

the result with the classical tools of time-series forecasting such as ARIMA and GARCH.

ACKNOWLEDGEMENTS

The author is so grateful to Assoc. Prof. Dr. Krisorn Jittorntrum for the Neural Networks and the intellectuality he taught and brought to the author.

REFERENCES

Gurney, K. 1999. An Introduction to Neural Networks. 2nd ed. London : UCL Press.

Haykin, S. 1994. Neural Networks : A Comprehensive Foundation. New York: Macmillan College Publishing Company.

Patterson, D.W. 1996. Artificial Neural Networks : Theory and Applications. Singapore : Prentice Hall.

Smith, K.A., and J.N.D. Gupta. 2000. Neural networks in business: techniques and applications for the operations researcher. Computers and Operational Research 27: 1023-1044.

none 386