

A Study on the Maintenance of Distributed Lag Model in Time Series Prediction Model

Jung-Ho Choo, Yu-Jin Kim, and Jung-Ho Eom(✉)

Daejeon University, 62 Daehakro, Dong-Gu, Daejeon 300-716, Republic of Korea
eomhun@gmail.com

Abstract. An ambiguity in time series prediction analysis is that the distributed lag model can be transformed due to AR, differencing, MA, and seasonality adjustment. This means that the time series prediction model can make predictions without considering distributed lag model, which is a characteristic of time series. This study is to confirm whether the time series prediction model maintains the distributed lag model. Analysis tools were used cross-correlation and Granger causality test. Time series forecast models were used statistics-based ARIMA model and RNN-based LSTM model. And we used monthly terror attack data and monthly time series data of international oil prices from 1996 to 2020. The research method is four stages. First, we performed cross-correlation and Granger causality tests. Second, the ARIMA model and LSTM model were applied to the time series. Thirdly, we reperformed cross-correlation and Granger causality tests on the predicted results. Fourth, in the conclusion, the first and third steps results were compared to confirm whether the prediction model maintained distributed lag model. As a result of this research method, the ARIMA model maintained the distributed lag model in the cross-correlation, and the LSTM model maintained the distributed lag model in the Granger causality test.

Keywords: Distributed lags model · ARIMA · LSTM · Cross-correlation · Granger causality

1 Introduction

This study is to confirm whether the forecasting model maintains the distributed lag model. Time series prediction analysis may not maintain the distributed lag model due to AR, differencing, MA, and seasonality adjustments. This means that time series prediction results may differ from original time series the distributed lag model. Differencing adjusts the trend of the time series and make it stationary. Therefore, the original time series and the Undifferentiated time series are different from each other. However, in the ARIMA model, the differencing equalizes average and dispersion in the time series so that the time series has stationarity. Differencing in LSTM (Long-Short-Term-Memory) models improve predictive accuracy. MA (Moving Average) or SMA (Seasonal Moving Average) can change the characteristics of distributed lag model. The SMA has a good effect of seasonality adjustment. However, if this is applied excessively, it is different from the appearance of the original time series. MA represents the current data as a

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14105, pp. 374–391, 2023.
https://doi.org/10.1007/978-3-031-37108-0_24

combination of white noise. This characteristic of MA allows effective analysis by mitigating non-uniform delay of time series data. However, this characteristic of MA broke the relationship if there was a causal relationship between other time series.

Therefore, we applied the ARIMA model and the LSTM model to the time series data of the current society, and confirmed whether the forecasting model maintains the distributed lag model. We selected Cross-correlation and Granger causality test as methods to determine whether the distributed lag model is maintained. We used the time series by monthly terror attack and international oil prices by month from 1996 to 2020. We empirically analyze the hypothesis that the time series prediction model maintains the distributed lag model using real society time series data. Under this hypothesis Sect. 2 reviews the related work to this study and explains the analysis tools and order of the study. In Sect. 3, time series data are analyzed in the order of research set in Sect. 2. Finally, the research hypothesis is evaluated with the results of Sect. 3.

2 Related Work and Method

2.1 Related Work

In Kim's paper [2], compensated the Granger causality test of the index according to economic policy with stock returns and gold futures. And this paper predicted the closing price of asset prices with a deep learning LSTM model. Through the Granger causality test, this paper was analyzed the impact of the economic policy index on the rate of return and money in bitcoin, predicted the public with an LSTM model, and analyzed the model's cooperation. As a result, it was confirmed that there is a significant relationship between the indicator and the increase in Bitcoin price. In addition, in the shock response function analysis, there was a supportive response to the economic policy index in future assets and finally, in the LSTM analysis, the digital self-prediction index was not repeated. The time series recording the KOSPI index represents an excellent predictor of economic policy indicators [2].

In Sun's paper [3], applied ARIMA and VAR models to portal and Internet information media service industries, information service industries, and online delivery industries. When being apply the information service industry to the ARIMA model, the most appropriate models are ARIMA $(1, 1, 1)(0, 1, 1)_{12}$ and ARIMA $(2, 1, 0)(1, 1, 0)_{12}$, it respectively showed an accuracy of 5.226% and 5.044%. The Granger causality test between the production index of database online providers and the portal and information-media service industries showed a significant result of 5%, and database online providers and portal and Internet information-media services had a significant effect on the wholesale and retail industries [3].

In Lee's papers [4], analyzed the causal relationship between two variables of housing price and transaction volume using the VEC model and compared the predictive power with the VAR model the VEC model and the AR model. For the forecast of time series, the out-of-sample prediction, dynamic, recursive forecast model method was applied to the time τ and then the prediction was made by increasing the period by 1 to the total $\tau + h$, and the accuracy was using Clark and McCracken's MES-F test and boot simulation. As a result of the study, there was a long-term positive correlation between transaction volume and housing price in the VEC model. But the causality from transaction volume

to price was low. In the comparison of forecast power was high in the order of the VEC model, VAR model and AR model [4].

According to Kim's research method, the Granger causality test between time series was performed and then prediction was made with the LSTM model, so it cannot be confirmed whether the causal relationship between time series after forecasting with the LSTM model is the same as before prediction. Sun's research method was to make time series predictions with an ARIMA model and then to confirm the causal relationship between time series. That is, it cannot be confirmed whether the causal relationship before applying the ARIMA model to the time series is the same as after forecasting. Lastly, Lee's research method is identified the causal relationship between time series with the VEC model, and then compared the predictive accuracy by predicting with the VAR model and the AR model. This research method confirmed the causal relationship between time series and then predicted. It is not possible to compare whether the causal relationship before and after the prediction is the same. What the three related work have in common is that they did not identify the causal relationship before and after prediction. In other words, they did not check whether the distributed lag model before and after the prediction was maintained [2–4].

2.2 Method

As mentioned earlier, the purpose of this study is to check whether the distributed lags model is maintained when applying prediction models to time series data. The source of the data was used GTD (Global Terrorism Database) [5] for terror attack time series data, and Korea National Oil Corporation (Petronet) [6] for international oil price time series data. And the program was used R version 4.2.1. The reason for using the terror attack time series is that it has seasonality. In other words, terror attack occurs uncertainly, so it may be disadvantages using time series analysis data. However, since there is a seasonality that occurs in a similar pattern every year, it can be used for time series analysis.

Standard analysis was performed using the Box-Jenkins method for ARIMA model analysis. As shown Fig. 1, the Box-Jenkins method proceeds in the order of identify model, parameter estimation, diagnostic model, model adoption and forecasting.

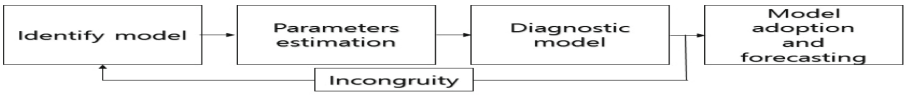

Fig. 1. Box-Jenkins method [7].

The ARIMA model consists of AR (Autoregressive), I (differencing) and MA (Moving average). The distinction between the ARIMA model and the ARMA model is the presence or absence of an I (differencing), which means that the ARIMA model can analyze time series data are non-stationary. The expression of AR is as follows.

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + e_t \tag{1}$$

The expression of I (differencing) is as follows.

$$1^{\text{st}}\text{differencing} : y_t^{'} = y_t - y_{t-1} \tag{2}$$

$$2^{\text{nd}}\text{differencing} : y_t^{''} = y_t^{'} - y_{t-1}^{'} \tag{3}$$

The expression of MA is as follows.

$$MA_t = \frac{1}{2k+1} \sum_{i=-k}^{k} Y_{t+k} \tag{4}$$

The LSTM model works as shown in Fig. 2. As noted earlier, the LSTM model is a deep learning analysis model that has improved RNN (Recurrent Neural Network). As shown in Fig. 2, we show that the problem of RNN, such as the slope loss problem, is improved by adding hidden layers and cell states to the LSTM model. In other words, the predictive power has improved as the partial learning information can be memorized and reflected in new learning.

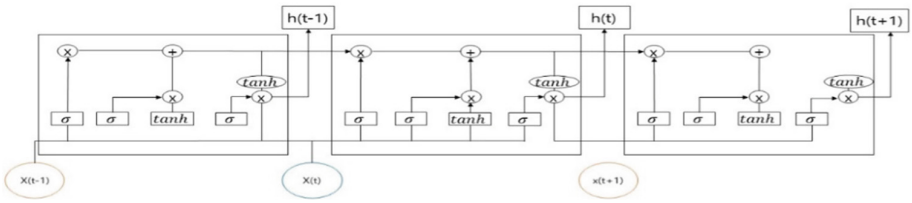

Fig. 2. LSTM architecture [8].

The LSTM model prevents the slope lose and effectively allows the slope to flow through three gates to protect and control the cell state the forget gate (f_t), input gate (i_t) and output gate. The forget gate (f_t) is a gate for forgetting past information. The output gate (0_t) of the sigmoid function is 0 to 1, if the value is 0, the information in the previous state is forgotten, and if the value is 1, the information in the previous state is completely remembered [9]. It remembers and receives input and cycles, takes s sigmoid, and sends it out [10]. The input gate (i_t) is a gate for storing flow information, and its sigmoid value is between 0 and 1 [9]. However, since $\tilde{c}_t$ is a Hadamard product and is the result of the hyperbolic tangent, its value is -1 to 1. That is, the value of sigmoid above may be -1 to 1, and the result may be a negative number [10]. Finally, the output gate (0_t) is the gate for the final results h_t, and the value obtained by Hadamard product of the hyperbolic tangent of the cell state is the final result of the LSTM. The formula for the LSTM cell is as follows [9].

$$f_t = \sigma_g\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{5}$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \tag{6}$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \tag{7}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \tag{8}$$

$$h_t = o_t \odot \sigma_h(c_t) \tag{9}$$

Cross-correlation is a way to check the similarity of two-time series [11]. This correlation analysis method is the correlation of time series observations through the delay between two-time series. Looking at the cross-correlation formula, we can see the similarity through the observations between the two-time series. The formula for the LSTM cell is as follows [11].

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(y_{k+k} - \bar{y})}{\sum_{t=1}^{N}(x_t - \bar{x})^2 \sum_{t=1}^{N}(y_t - \bar{y})^2} \tag{10}$$

Granger causality test for time series the standard causal relationship test method is different from correlation analysis because it identifies the causal relationship through a distributed lags model between time series. Therefore, this test method can be confirmed the direction of the causal relationship between time series. The formula for the Granger causality test is as follows [12].

$$Y_t = \sum_{i=0}^{m} \alpha_i X_{t-i} + \sum_{j=0}^{m} \beta_j Y_{t-j} + \epsilon_{1t} \tag{11}$$

$$X_t = \sum_{i=0}^{m} \lambda_i X_{t-k} + \sum_{j=0}^{m} \delta_j Y_{t-j} + v_{2t} \tag{12}$$

$$F = \frac{(SSE_R - SSE_{UR})/q}{(SSE_{UR})/(n-k)} \tag{13}$$

In the above F Eq. (13), SSE_R and SSE_{UR} are obtained as a sum of the two Eqs. (13). Subsequently the SSE is obtained when the constraint $\alpha_i = 0$, are given in (11) and $\delta_j = 0$ in (12) and when they are not. After that, if the value of the test statistics is greater than the threshold, the null hypothesis $H_0 : \alpha_i = 0$ or $H_0 : \delta_j = 0$ is rejected as the influence of the given condition is large.

This study is used to ARIMA, LSTM, cross-correlation, and Granger causality establish the hypothesis, method, and sequence of the study as follows.

H_o: Even if the prediction model is applied to time series data, the distributed lags model is maintained.

1. Granger causality test and Cross correlation of monthly terror attacks and international oil price from 1996 to 2020.
2. By forecasting three years of terror attacks and international oil price time series using the ARIMA model and LSTM model and generate time series data reflecting the forecast results.
3. Using two-time series data reflecting the forecasting results, the Granger causality test and Cross correlation before the prediction was performed.
4. Analyze the first and third results to evaluate the research hypothesis.

3 Analysis Result

3.1 Checking the Distributed Lag Model of the Original Time Series

The research method of the Granger causality test followed two steps. First, it is determined whether the first time series data is a stationary time series. The reason why time series data is determined to be stationary is that the variance of the mean increases with time lag in the case of non-stationary time series. If the Granger causality test is performed with a time series that does not have stationarity, may occur spurious regression results. To prevent this, an ACF (Auto-correlation function), PACF (Partial Auto-correlation function) and ADF (Augmented Dickey-Fuller) test was performed.

Second, it is to confirm the lag of the appropriate Granger causality test. The result of the Granger causality test is sensitive to the length of the selected lag. Therefore, in order to proceed smoothly with the Granger causal relationship, an appropriate number of lags must be determined. Since the data used are monthly terror attack and international oil price time series data, 1, 4, 6, 12, 18, 24… were set to reflect the level of interest by quarter and year.

Before proceeding with the Cross-correlation and Granger causality test between monthly terror attack and international oil prices from 1996 to 2020, it is necessary to determine whether the time series has stationarity. Figure 3 is a graph visualized to intuitively analyze the two-time series.

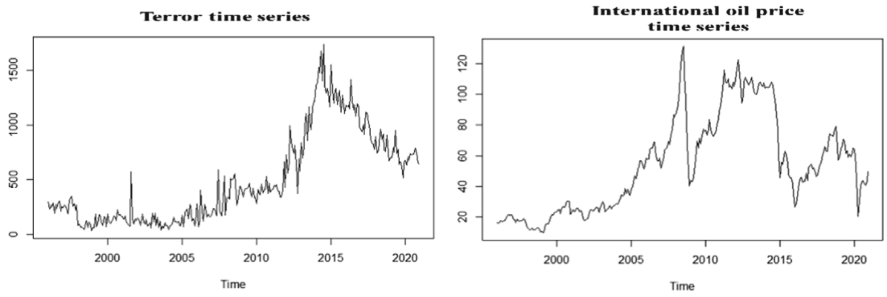

Fig. 3. Terror and international oil price General graphs

At the Fig. 3 of monthly terror attack and international oil prices from 1996 to 2020, both time series have points of sharp rise and fall, so we can suspect that the average may not be constant.

In addition, looking at the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) of terror attack and international oil prices in Fig. 4, there is a strong trend with a fast decrease as the time lag increases. That is, we assumed a time series non-stationarity.

It is necessary to check whether the two-time series data are stationary time series through the ADF test, which is a traditional stationarity confirmation method. ADF test results-Name of the terror attack time series is tr, and the international oil price is abbreviated to di-are shown in Table 1.

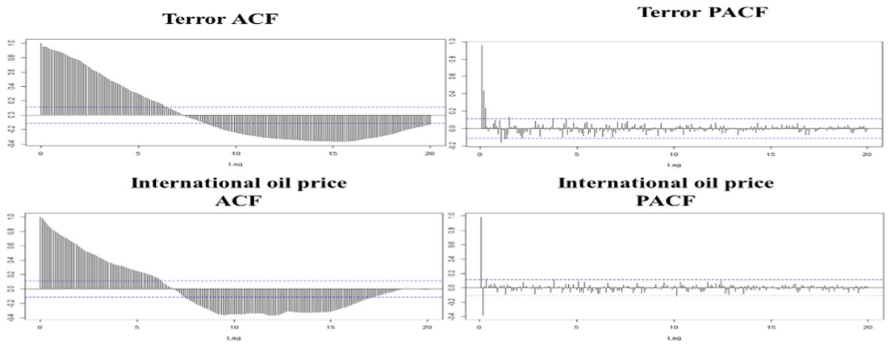

Fig. 4. ACF and PACF of terror attack and international oil prices time series.

Table 1. Result of ADF test.

ADF test	Dickey-Fuller	Lag order	p-value
tr	−1.7251	6	0.6926
tr, $d = 1$	−7.6349	6	0.01*
di	−1.8071	6	0.6576
di, $d = 1$	−7.103	6	0.01*

* in Table 1 is statistically significant as lower than 0.01. According to the ADF test results when the original time series and the 1st differencing of two-time series are performed, it can be seen that 1st differencing between the two-time series is required to become stationary time series data. So according to the results of the ADF test, the Granger causality test was used time series with the 1st differencing, and the graph of the series to which the 1st differencing was applied is shown in Fig. 5.

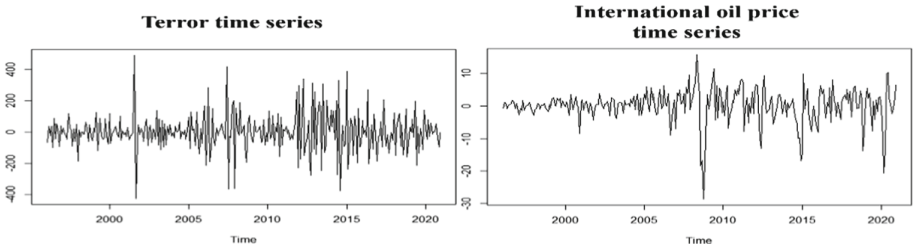

Fig. 5. General graph with 1st differencing applied two-time series.

Prior to the Granger causality test of the two-time series, cross-correlation analysis was performed to confirm the similarity between the time series. The reason for performing the cross-correlation analysis is that it analyzes the correlation between different time series between −1 and +1, so that the distributed lags models between time series can be compared simultaneously. Looking at Fig. 6, the highest cross-correlation is −

18 lag. So, we assumed that the terror attack time series can affect the international oil price time series with an 18 month frequency.

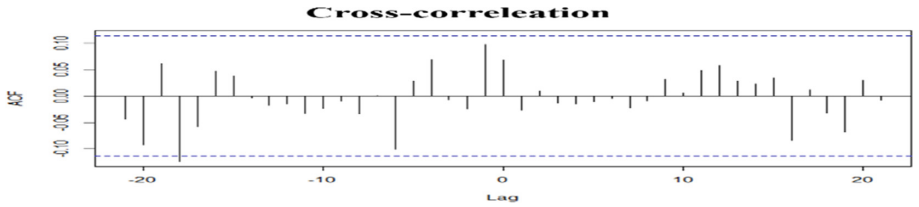

Fig. 6. Result of cross-correlation.

Cross-correlation used a distributed lag model between time series. In other words, cross-correlation is more difficult to confirm a clear causal relationship than Granger causality test. Therefore, in order to identify a clear causal relationship, a Granger causality test was performed between the two-time series, and the results are shown in Table 2.

Table 2. Result of Granger causality test.

Lag	tr → di		di → tr	
	F-statistic	P-value	F-statistic	P-value
24	1.4323	0.0935*	0.8374	0.6868
30	1.4002	0.098085*	1.0303	0.4297
84	0.7928	0.8228	1.6049	0.4064*

* in Table 2 is statistically significant as lower than 0.05. Table 2 shows the results of the Granger causality test between terror attack and international oil price time series. In the causal relationship direction of terror attack in international oil prices, it can be confirmed that there is a causal relationship with a significant probability of less than 5% of in the lag of 24 and 30. Conversely in the causal relationship from international oil prices to terror attack, it was confirmed that there was a significant causal relationship with a probability of less than 5% of in the lag of 84. Now we have performed cross-correlation and Granger causality test as time series of the original state. Therefore, after forecasting three years with the ARIMA model and LSTM model and then cross-correlation and Granger causality test should be performed again.

3.2 ARIMA Model

Before forecasting a time series with an ARIMA model, we verified that original time series are not stationary in previous Sect. 3.1. Therefore, in the ARIMA model analysis, we did beyond the process of identify model and proceed with the method of parameters estimation of Box-Jenkins method.

In the parameters estimation stage, ARIMA $(0, 1, 1)(0, 0, 2)_{12}$ was selected as an appropriate ARIMA model for time series in the case of terror attack. And ARIMA

$(0, 1, 1)(0, 0, 2)_{12}$ reflecting the seasonality existing in the terror attack time series is analyzed as the SARIMA model. In the case of international oil prices, seasonality did not exist. So it was analyzed as an ARIMA model, and ARIMA(0,1,2) was selected as an appropriate model. And the results and model values of these two models are shown in Table 3.

Table 3. ARIMA model values.

Model	ARIMA $(0, 1, 1)(0, 0, 2)_{12}$	ARIMA(0,1,2)
AIC	3607.59	1751.34
σ^2	9873	20.05
Ljung-box test	Pass	Pass

When forecasting three years with the ARIMA $(0, 1, 1)(0, 0, 2)_{12}$ model and the ARIMA(0,1,2) model the diagnostic values in Table 4. The forecasted graph in Fig. 7 are shown.

Table 4. ARIMA model forecast diagnostic values.

Diagnostic	ARIMA $(0, 1, 1)(0, 0, 2)_{12}$	ARIMA(0,1,2)
MAPE	24.89029	6.702753
RMSE	99.19756	4.470594

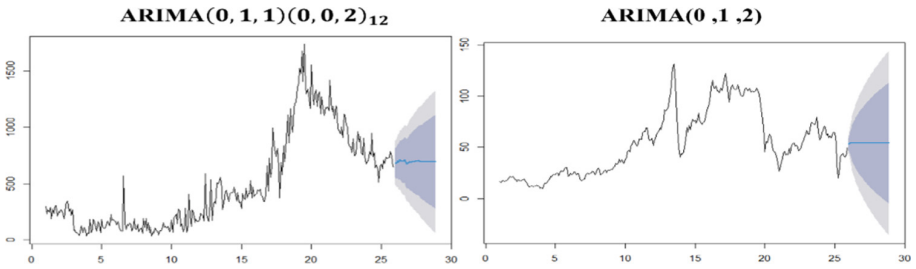

Fig. 7. ARIMA model Forecast graph.

Table 4 shows the diagnostic values of the ARIMA model. In the case of terror attack, seasonality exists, so it is assumed that MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Squared Error) values tend to be high. In the case of international oil prices, it was assumed that the values of MAPE and RMSE tended to be low because there was no seasonality.

3.3 LSTM Model

The LSTM model analysis will follow the operation of Fig. 2. And before applying the time series as the LSTM model, we willperform a 1st differencing to improve the model forecast performance and adjust with a stationary time series. Since the 1st differencing was made between the two-time series in the Granger causality test, the time series with the 1st differencing can be referred to Fig. 5.

As previous noted, the LSTM model is a technique for forecasting the future by learning the time series. So, when we analysis the time series with the LSTM model, there is something to be careful. It is to set the model learning rate and the layer dropout. In this study, we set the learning rate of 80% for both time series so that it can be forecasted after sufficient learning.

And the value of layer dropout with a learning rate of 80% was predicted for 3 years with the value when the lowest over fitting was confirmed. The above verification method confirmed the prediction accuracy by RMSE.

Table 5. LSTM model diagnostic values.

Diagnostic	terror attack	International oil prices
RMSE	95.21628	5.159664

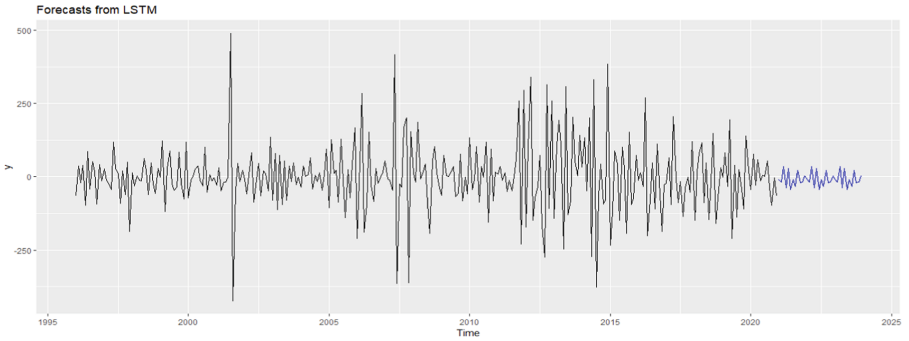

Fig. 8. Prediction results of the terror attack using LSTM model.

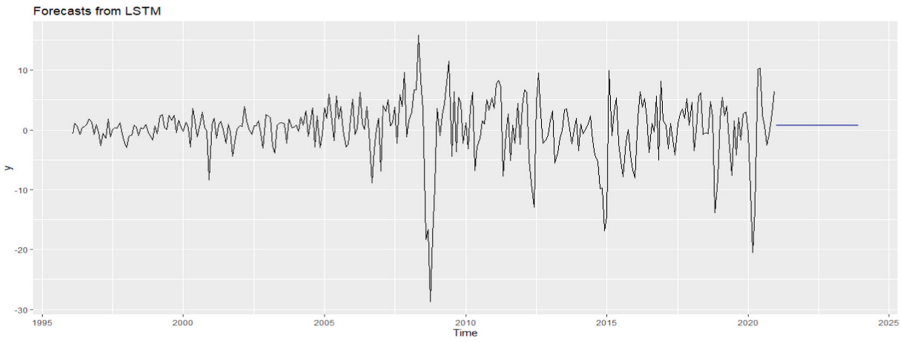

Fig. 9. Prediction results of the international oil prices using LSTM model.

Table 5, Fig. 8, and Fig. 9 show the RMSE and forecasting result when 80% of the learning rate and the layer dropout value to minimize overfitting are applied to each time series. In the case of terror attack, when the learning rate was 80% the layer dropout to minimize overfitting was 0.4% and the RMSE at this rate was 95.21628. In the case of international oil prices, the value of layer dropout, which minimizes overfitting was 0.1 and the value of RMSE was 5.159664.

An interesting point in the LSTM model analysis of terror attack and international oil prices is that, like the ARIMA model analysis, in the case of international oil without seasonality, the predicted values are static. In addition, the LSTM model was superior in the prediction model diagnosis value of RMSE, the terror attack time series prediction, and the ARIMA model was superior in the international oil price time series. In addition, in the prediction model diagnosis value of RMSE, the LSTM model was superior in the terror attack time series prediction and the ARIMA model was superior in the international oil price time series.

3.4 Checking the Distributed Lag Model of the Time Series with Forecast Model Applied

We obtain the time series data to which the prediction results through the analysis of the ARIMA model and the LSTM model. So, we could confirm whether the time series to which the prediction results are applied maintains the distributed lag model of the original time series. In this table of contents, we performed the Granger causality test by confirming the quality of the time series and applying the differencing according to the stationary of the time series, like the previous research method of the Granger causality test. We set the lag of the Granger causality test to 1, 4, 6, 12, 24…, and the cross-correlation to a lag of −21 to 21 for the original time series. The time series with predictive model applied had more observations.

Before analyzing cross-correlation and the Granger causality test, we confirmed the stationary of the time series to which the forecasted results of the ARIMA model and the LSTM model were used by the ACF, PACF, ADF test. It was shown in Fig. 10, Fig. 11, and Table 6.

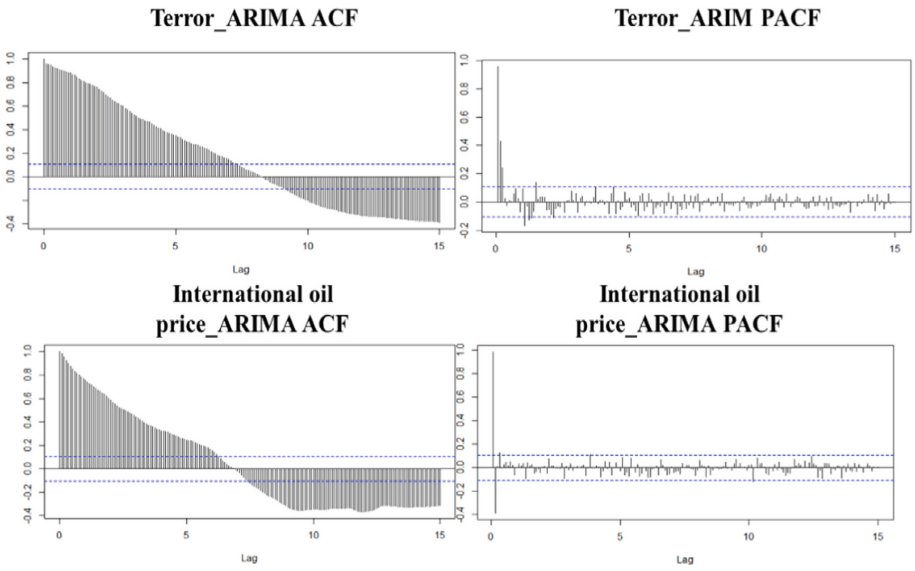

Fig. 10. ACF and PACF of ARIMA model.

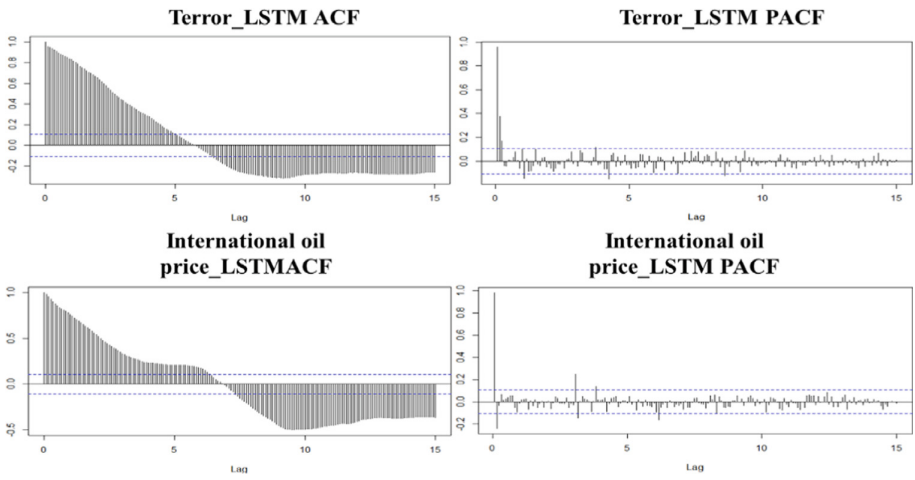

Fig.11. ACF and PACF of LSTM model.

Looking at Fig. 10 and 11, the ACF and PACF of both models decrease rapidly with time. This phenomenon is that both the ARIMA model and the LSTM applied time series have a strong trend, so they can't be time series that do have stationarity. That is, it is necessary to determine whether the time series is stationary through the ADF test.

Table 6. ADF test of applied time series with prediction results.

ADF test	Dickey-Fuller	Lag order	P-value
tr_ARIMA	−1.6464	6	0.7528
di_ARIMA	−1.91	6	0.6146
tr_ARIMA, $d = 1$	−8.1436	6	0.01*
di_ARIMA, $d = 1$	−7.5691	6	0.01*
tr_LSTM	−0.94363	6	0.9467
di_LSTM	−1.2007	6	0.9053
tr_LSTM, $d = 1$	−7.7046	6	0.01*
di_LSTM, $d = 1$	−8.1432	6	0.01*

* in Table 6 is statistically significant as lower than 0.01. We confirmed the stationary of the time series with the ARIMA model and LSTM model applied were judged to be times series without non-stationary when $d = 0$.Therefore, in order to analyze in a time series having stationary, the first order differencing had to be applied. Figure 12 shows the general graph of the time series in which the time series with stationary was found by the ADF test.

Prior to the Granger causality test, the distributed lags model of the times series was identified through cross-correlation analysis that can find out the similarity between the two-time series. In the case of the time series with the ARIMA model applied on the left of Fig. 13, the two-time series are similar with a lag of −18, and then similar with a lag of −1. In the case of the time series with the LSTM model applied on the right side of Fig. 13, the highest similarity was observed at −20 lag, followed by the highest similarity at 0 lag. We compared the cross-correlation function of the original time series data shown in Fig. 6 with the cross-correlation function of the time series with the prediction model applied. In the case of the ARIMA model, the two-time series at −18 lag, the same lag as the original time series, and −1, the next similar lag. The result is similar, so that the distributed lags model was considered when the ARIMA model performed the prediction, and the difference in cross-correlation values with the original time series was an average of 1.1%. So, in the case of ARIMA model we can support the research hypothesis H_0 that we established earlier.

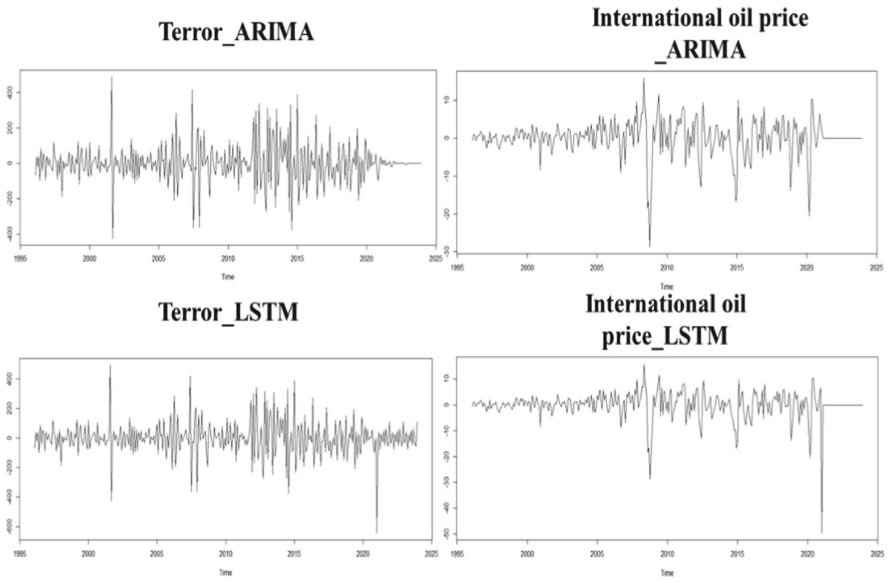

Fig. 12. 1st differencing graph of time series data with forecast results applied.

On the other hand, in the case of the time series with the LSTM model applied, it hasn't a similar distributed lags model to the original time series because it shows high similarity at lags of −20 and 0. And the difference in cross-correlation values has a large difference of about 20% on average. According to the result of the cross-correlation function, the ARIMA model performs prediction considering the distributed lags model of the original time series, and it is insufficient to support H_0 in the cross-correlation of the LSTM model prediction.

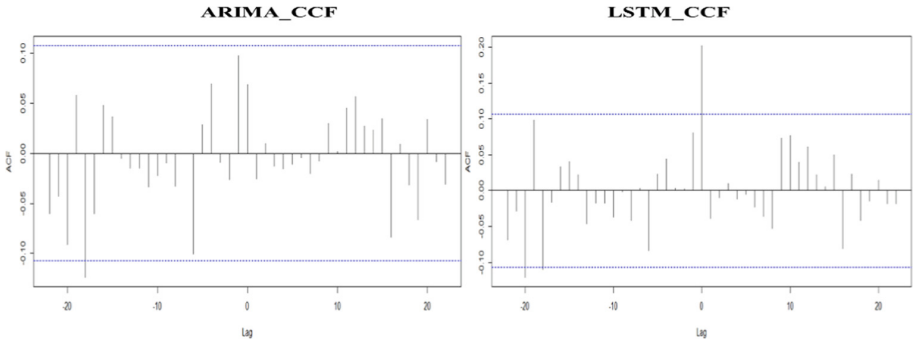

Fig. 13. Applied cross-correlation of ARIM model and LSTM model.

When the Granger causality test of the time series data to which the prediction results of the ARIMA model are applied is performed, the results shown in Table 7. * in Table 7 is statistically significant as lower than 0.01.

Table 7. Granger causality test results of ARIMA model applied time series.

Lag	tr_ARIMA → di_ARIMA		di_ARIMA → tr_ARIMA	
	F-statistic	P-value	F-statistic	P-value
6	0.1823	0.9816*	2.1657	0.04608*
18	0.4958	0.9589*	1.4844	0.09441*
24	0.9087	0.5903	1.5805	0.04493*
30	1.1389	0.29	1.4857	0.05595*
36	1.259	0.1607	1.3784	0.08508*
48	1.3495	0.0819*	1.116	0.2981
54	1.4906	0.02851*	0.9013	0.6653
60	1.3581	0.0694*	0.9827	0.5195
72	1.3488	0.07457*	0.9058	0.6727
78	1.4165	0.05065*	0.9267	0.6352
84	1.6718	0.01022*	0.8987	0.6865
90	1.4826	0.0484*	1.1459	0.2838

* in Table 7 is statistically significant as lower than 0.05. Looking at the results of the Granger causality test of the time series with the ARIMA model applied in Table 7, the results contradicted the results of the Granger causality test of the original time series data. In the Granger causality test of the original time series data, the terror attack variable affects the international oil price at lags of 24 and 30, and that the international oil price affects terror attack at the lag of 84. However, the Granger causality test results of the time series to which the prediction results of the ARIMA model were applied showed that terror attack affects international oil prices at lags of 48, 54, 60, 72, 78, 84, and 90. International oil prices were found to have an effect on terror attack at time lags of 6, 18, 24, 30 and 36. In other words, the causality of the two-time series changed by being apply the ARIMA model. This can be said that the causality of the original time series to which the prediction results of the ARIMA model are applied has changed, and unlike the cross-correlation results, it is difficult to support H_0.

Table 8. Granger causality test results of LSTM model applied time series.

Lag	tr_LSTM → di_LSTM		di_LSTM → tr_LSTM	
	F-statistic	P-value	F-statistic	P-value
78	1.3152	0.09803*	1.0123	0.4736
84	1.4893	0.03588*	0.9293	0.46308
102	1.3615	0.176	1.6516	0.06452*

* in Table 8 is statistically significant as lower than 0.05. The Granger causality test result of the time series to which the prediction result of the LSTM model is applied, it is similar to the Granger causality test of the original time series. According to Table 7, in the causal direction of international oil prices in terror attack, causality exists at lags of 78 and 84. Conversely, in the causal direction of terror attack in international oil prices, causality exists at a lag of 102. It can be judged that the LSTM model maintained the distributed lag model. Because the lag occurred at an interval of 6 in the causal direction of the international oil price in the original time series terror attack. Also, in the causal direction of terrorism in international oil prices, causality appeared at a lag of 84 in the original time series, and a time series was formed at 102 in the time series to which the LSTM model was applied. These results it that predicting the LSTM model considers the distributed lag model of the time series used for prediction. Because there is the characteristic that the Granger causal test uses the distributed lag model. That is, in the cross-correlation function analysis, the ARIMA model considers the lag distribution model when making predictions, and in the Granger causal test, the LSTM model considers the lag distribution model when making predictions, and the H_o hypothesis can be supported.

4 Conclusion

In this study, terror attack and international oil price time series data were used with ARIMA, a statistical prediction model, and LSTM, a deep learning prediction model. We confirmed whether these models maintain the distributed lag model when these models predict through cross-correlation and Granger causality tests. The results of the Granger causality test when the prediction model was not applied had causality at 24 lags and 30 lags in the causal direction of international oil prices in terror attack. In the causal direction of terror attack in international oil prices, causality was confirmed at 84 lags. When the ARIMA model was analyzed through the Box-Jenkins method, the terror attack time series data that is reflected the seasonality was judged to be suitable the SARIMA model of ARIMA $(0, 1, 1)(0, 0, 2)_{12}$ as MAPE 24.89029 and RMSE 99.19756. Since the ARIMA analysis of international oil prices has no seasonality, ARIMA (0,1,2) was judged to be an appropriate model with MAPE 6.702753 and RMSE 4.470594. When the LSTM model was analyzed with a learning rate amount of 80%, the appropriate layer dropout value for the terror attack time series data was 0.4, and the RMSE at this time was 95.21628. The appropriate layer dropout for international oil prices was 0.1, at which time the RMSE was 5.159664. When the prediction results obtained by the ARIMA model and the LSTM model were applied to the original time series, we reperformed the cross-correlation analysis and the Granger causality test. The results is: in the case of the ARIMA model, the two-time series were most similar at -18 lag which is a very similar result to the cross-correlation analysis of the original time series,

and then similar at −1 lag. In addition, it was judged that the accuracy of the average value of cross-correlation differed by about 1.1%, it was predicted similarly to the lag distribution model of the original time series. Conversely, the cross-correlation analysis of the LSTM prediction model showed that the two-time series were most similar at − 20 lag, followed by 0 lag. Numerically, there was a difference of about 20.8% from the original time series. It was judged that the ARIMA model maintains the distributed lag model and forecasts rather than the LSTM model, even considering that the result of the cross-correlation increases the observed value when the forecasting model is applied to the time series.

On the contrary, in the Granger causality test, causality was confirmed at lags of 48, 54, 60, 72, 78, 84, and 90 in the causal direction of international oil prices in terror attack in the predicted time series data of the ARIMA model. In the causal direction of terror attack in international oil prices, causality was confirmed at lags of 6, 18, 24, and 30, which is quite different from previous causal relationship tests between existing time series. As a result of the prediction of the LSTM model, time series data was confirmed causality at lags of 78 and 84 in the causal direction of international oil prices in terror attack. In the causal direction of terror attack in international oil prices, causality was confirmed at a lag of 102. We compared the Granger causality test results of the original time series and the test results of the time series to which the forecasting model was applied. Considering that the observed increased by applying the forecasting model, the result of the causal direction of international oil prices in terror attack was 78, 84 lags. It was the same as that at an interval of 6, like the original time series. In the causal direction of terror attack in international oil prices, it can be judged that only one causality was confirmed, such as 84 and 102, respectively, maintaining the distributed lag model of the original time series data. In other words, in the cross-correlation analysis, the ARIMA model originally maintained the distributed lag model of the time series data, and in the Granger causal test, the LSTM model maintained the original distributed lag model of the time series data.

As mentioned above, the purpose of this study was to confirm whether the forecasting model maintains the distributed lag model when making predictions. The analysis tools used to confirm this were the Granger causality test, which confirms causality between time series with a distributed lag model, and the cross-correlation, which confirms the similarity of two-time series through lag. Since the Granger causal test and cross-correlation both use the distributed lag model, the distributed lag model of the time series to which the original time series and forecasting model was applied was confirmed. But there is a limitation that it was not compared with an accurate distributed lag model of the time series. In future research, we will devise a research method that can accurately observe this distributed lag model so that we can get closer to the essence of the predictive model. In addition, this study applied ARIMA model and LSTM model to specific time series, such as terror attack and international oil prices, it is unknown whether the distributed lag model is maintained when the forecasting model is applied to other time series. In the next study, we will try to devise whether other prediction models other than the ARIMA model and the LSTM model maintain the distributed lag model, and how the forecasting model can effectively maintain the distributed lag model.

Acknowledgement. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5C2A03093531).

References

1. Choi, Y.W.: Accuracy improvement technique for the prediction of future living population based on the Diff LSTM algorithm. Graduate School of Information and Communications of Sungkyunkwan University, pp. 1–50 (2022)
2. Kim, S.W.: Prediction performance of LSTM-based economic policy uncertainty index on asset prices: traditional asset vs digital asset. J. Digit. Contents Soc. **23**(6), 1105–1113 (2022)
3. Sun, I.-S., Park, S.H.: Prediction of information service industry & analysis of its causal relation with wholesale & retail industries using time series analysis. e-Bus. Stud. **15**(6), 101–120 (2014)
4. Lee, Y.S., Lee, J.P.: Causality and predictability of price-volume in housing market-evidence from Seoul and Busan housing markets. SH Urban Res. Insight **8**(3), 51–67 (2018)
5. GTD Hompage. https://www.start.umd.edu/gtd/. Accessed 09 Jan 2023
6. Petronet Homepage. https://www.petronet.co.kr/v3/index.jsp. Accessed 09 Jan 2023
7. Tang, Z., De Almeida, C., Fishwick, P.A.: Time series forecasting using neural networks vs. Box-Jenkins methodology. SIMULATION **57**(5), 303–310 (1991)
8. Dudek, G., Pełka, P., Smyl, S.: A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting. IEEE Trans. Neural Netw. Learn. Syst. **33**(7), 2879–2891 (2021)
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
10. Nifa, K., Boudhar, A., Ouatiki, H., Elyoussfi, H., Bargam, B., Chehbouni, A.: Deep learning approach with LSTM for daily streamflow prediction in a semi-arid area: a case study of Oum Er-Rbia river basin, Morocco. Water **15**(2), 262 (2023)
11. Bourke, P.: Cross correlation. In: Cross Correlation, Auto Correlation—2D Pattern Identification (1996)
12. Hur, Y.K., Jang, K.S., Kim, S.J., Kim, H.M.: The granger casuality analysis between prices and trading volume in the housing market-focused on apartment property markets in Seoul. Hous. Stud. Rev. **16**(4), 49–70 (2008)