*International Journal of Advanced Smart Convergence Vol.12 No.4 75-87 (2023)*
*http://dx.doi.org/10.7236/IJASC.2023.12.4.75*

IJASC 23-4-9

# Development of a Model to Predict the Volatility of Housing Prices Using Artificial Intelligence

[1]Jeonghyun LEE, [2]Sangwon LEE

*[1]Researcher, Department of Real Estate Studies, Konkuk University, Korea*
*leejohn.news@gmail.com*
*[2]Prof., Dept. of Computer & Software Engineering, Wonkwang Univ., Korea*
*sangwonlee@wku.ac.kr*

***Abstract***

*We designed to employ an Artificial Intelligence learning model to predict real estate prices and determine the reasons behind their changes, with the goal of using the results as a guide for policy. Numerous studies have already been conducted in an effort to develop a real estate price prediction model. The price prediction power of conventional time series analysis techniques (such as the widely-used ARIMA and VAR models for univariate time series analysis) and the more recently-discussed LSTM techniques is compared and analyzed in this study in order to forecast real estate prices. There is currently a period of rising volatility in the real estate market as a result of both internal and external factors. Predicting the movement of real estate values during times of heightened volatility is more challenging than it is during times of persistent general trends. According to the real estate market cycle, this study focuses on the three times of extreme volatility. It was established that the LSTM, VAR, and ARIMA models have strong predictive capacity by successfully forecasting the trading price index during a period of unusually high volatility. We explores potential synergies between the hybrid artificial intelligence learning model and the conventional statistical prediction model.*

***Keywords:*** *Artificial Intelligent, Bigdata, Housing Price, Prediction, Volatility.*

## 1. Introduction

In Korean society, a family's 'house' comprises the majority of its possessions. A family is greatly impacted by changes in real estate values. There has long been a belief that real estate is a stable asset. Politics, the economy, and industry are just a few of the variables that have an impact on the real estate market. Because of this, it is exceedingly hard to forecast changes in price. This study's primary goal is to employ Artificial Intelligence [1, 2, 3, 4] technologies with Bigdata [5, 6, 7, 8] to forecast changes in real estate values and identify their underlying causes. For a considerable amount of time, econometric-based time series analysis has been used for price prediction. There were advantages and disadvantages to the econometrics time series analysis model. There have been attempts to apply research findings from the engineering community to real

Manuscript Received: october. 10, 2023 / Revised: october. 15, 2023 / Accepted: october. 20, 2023
Corresponding Author: Sangwon LEE (sangwonlee@wku.ac.kr)
Tel: +82-63-850-6566
Author's affiliation: Prof., Dept. of Computer & Software Engineering, Wonkwang Univ., Korea

estate studies as artificial intelligence research has advanced to a full-fledged state. Academic real estate has long attempted to incorporate artificial intelligence techniques into time series analysis methods based on available statistics. Numerous people believed that artificial intelligence-based price prediction systems would be able to surpass the constraints of current time series analysis and forecast future real estate values with greater accuracy. In real estate studies, real estate price prediction has long been a significant area of study. Time series analysis using econometrics forms a significant portion of the prediction method. The aim of this research is to determine how to leverage the benefits of recently identified artificial intelligence methods while retaining the long-term research-verified benefits of the econometric model. The price prediction techniques based on artificial intelligence and metering economy are first verified for this research project in order to ascertain the advantages and disadvantages of each. After that, we figure out how to combine and utilize the two techniques. Following the confirmation of the prediction's accuracy, we examine the process used to identify the features of the real estate market for every time period. Various factors occasionally have an impact on the real estate market. It is true that a variety of political and economic factors influence prices in addition to changes in interest rates. The methodical identification of variables that influence the real estate market at a given moment and the development of appropriate real estate policies are two more significant goals of this research.

## 2. Related works

### 2.1. ARIMA model

Time series analysis attempts to identify characteristics only with its own current and past values without considering other variables in causal relationship with the characteristics value when analyzing them. General regression analysis does not take into account timing, but time series prediction seeks to create a statistically valid model for the purpose of collecting and analyzing historical observations for predictors and predicting the future. Auto-Regressive Integrated Moving Average (ARIMA) [9] Model, a linear regression-based univariate time series prediction model, generalizes the Auto-Regressive Moving Average (ARMA), a methodology for predicting the future through errors with past observations, and generalizes the trend of past observations by including the trend of past observations in Box-Jenkins' algorithm that combines the Auto-Regression (AR) and Moving Average (MA). The three stages of Box-Jenkins' model fit for analyzing time series data using ARIMA consist of predicting using the finally selected model after identifying, estimating, and diagnosing the model. In the model identification stage, the normality of the time series is first reviewed and then identified by referring to the forms such as Authorization Function (ACF), Partial Auto Correlation Function (PACF), and Inverse Auto Correlation Function (IACF).

### 2.2. VAR model

After Sims first proposed the Vector Auto-Regression (VAR) [10] model, it was technically modified and supplemented by Watson. It is a method of analyzing the results derived from actual data rather than economic theory. The VAR model assumes a simple structure with a multivariate autocorrelation model, but it can flexibly model the autocorrelation structure of target variables than univariate autocorrelation, and the dynamic response of a variable's change to endogenous variables can be identified through impact response analysis.

### 2.3. LSTM model

Long Short-Term Memory (LSTM) [11] model is a model suitable for long-term time series data prediction and is a widely used technique for time series data prediction. LSTM has been widely used to predict existing stock prices. Although these LSTM techniques have already been used a lot in predicting stock market prices, they have not been applied to the real estate market. Neural networks have been evaluated to be useful for predicting single time series data. For this reason, existing research has steadily developed. Recurrent Neural Networks (RNN), Echo State Networks (ESN), Generalized Regression Neural Networks (GRNN), LSTM are its examples. In particular, RNN, a deep learning technique, are widely applied to time series data prediction. RNN was developed to overcome the limitation of placing only neurons without considering context in the hidden layer for artificial neural network analysis. RNN is composed of a directed cycle in which past data can affect future prediction results. An RNN that processes sequential information is suitable for processing time series data. Considering the correlation between the previous data (t-1) and the current data (t), it is a neural network model that considers the past data that can predict the data of the next (future) t+1. However, RNN has a Vanishing Gradient problem over time, and LSTM supplements this. This has been studied in a way that is appropriate for predicting the future considering past data macroscopically.

## 3. Model to predict the volatility of housing prices

In this study, variables that can affect real estate sales prices and real estate prices from February 2005 to September 2019 were collected through the Bank of Korea's Economic Statistics System (ECOS) and the Korea Real Estate Agency's Real Estate Statistics Information System (R-ONE). For comparative analysis with ARIMA, VAR, and LSTM algorithms, data is collected and processed in three cycles based on the period of increasing volatility described above. It was defined as shown in Table 1 according to the period. Data for each cycle period were set as the training period and the volatility increase period was the test period for the period maintained by a certain trend.

**Table 1. Setting experimental periods**

| Type | Experiment | Period |
|---|---|---|
| The first cycle | Exp1 | Training period (43): 2005.02. ~ 2008.09.<br>Test period (6): 2008.10. ~ 2009.03. |
| The second cycle | Exp2 | Training period (36): 2009.04. ~ 2012.03.<br>Test period (17): 2012.04. ~ 2013.08. |
| The third cycle | Exp3 | Training period (55): 2013.09. ~ 2018.03.<br>Test period (18): 2018.04. ~ 2019.09. |

### 3.1. ARIMA data

Since ARIMA is an auto-regressive cumulative movement average model, analysis was performed using only the trading price index variable, and a normality test was performed to analyze using ARIMA. Considering the fact that it is an analysis using time series data, the stability of the time series data was first checked. To this end, a unit root test was performed, and the test method used the Augmented Dickey-Fuller (ADF) test method. Unit root test is performed and stable time series data is made after difference for each trading price index. We would like to conduct an analysis using differentiated data for each region. Through the difference, the normality condition is satisfied for each home sales price index and proceeds. After setting the difference (d) for each region, experiments are needed in various combinations for each home sales price index that determines p and q. In this paper, AICc is calculated for various combinations through a Grid Search method that lattice-tunes p and q, such as ARIMA hyperparameter tuning (p, q), and p and q with the lowest

AICc are selected (in Figure 1). Through various experiments, parameter tuning is performed, and the model with the lowest AICc is selected for each region.

$$AICc = -2\ln(L) + \frac{2(p+q+1)n}{n-(p+q)}$$

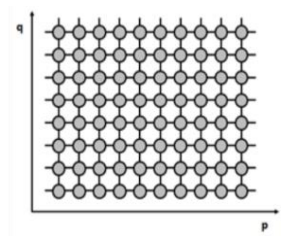


**Figure 1. AICc with p and q**

### 3.2. VAR data

In this study, variables that can affect real estate sales prices and real estate prices from February 2005 to September 2019 were collected through the Bank of Korea Economic Statistics System (ECOS) and the Korea Real Estate Agency Real Estate Statistics Information System (R-ONE). Changes in the sales price index (Apartments) across the country were first collected during the period. Various variables that can affect prices were collected from the Bank of Korea's economic statistics system. The variables collected include corporate bond yield, consumer price index, M2 (currency volume), mining industry index, and lease price index (KB). These variables are divided into the same training and test period as ARIAM to verify predictive performance.

### 3.3. LSTM data

In the LSTM, for the safety of learning, the source data is normalized through some data preprocessing processes. By unifying each data scale from 0 to 1, the instability of learning due to values is reduced. The input data structure for implementing the LSTM model for predicting the price is a three-dimensional structure of Batch Size, Time Steps, and input lengths, and three batches, sequence length, and input dimension.

* Y(t) = f(X1,X2,X3,X4,X5,X6)
* f: ARIMA or LSTM or VAR

X1: M2 (average balance, seasonal adjustment series)
X2: Production and consignment inventory index by the mining industry
X3: Consumer price index
X4: Housing price index (KB)
X5: Corporate bond yields
X6: Sales price index

* Y(t) = f(X6)
* f: ARIMA

**Figure 2. Basic data configuration**

### 3.4. Configuring scenario and datasets

The function of the five derivative variables selected and the home sales price index is shown in Figure 2. With each trading price index as a dependent variable (Y), the found derivative variables are composed of explanatory variables and analyzed. Data was organized to predict the trading price index. It was selected as an analysis period from February 2005 to September 2019, and was divided into training data and test data. The training data set is called a textbook, and the test data set is called a university entrance exam.

# 4. Experiment and its results

## 4.1. Performance comparison

It compares the performance (MAE, RMSE) for ARIAM, LSTM, and VAR of the national trading price index with accuracy (3%, 5%, 10%). October 2008, April 2012, and April 2018 were periods when the national sales price index (apartments) fell, resulting in increased volatility in the real estate market. In this situation, the change in the trading price index is predicted. Figure 3 shows the visualization of the national data set, and the index decline period of Exp1 was 6 months, which was shorter than 17 months of Exp2 and 18 months of Exp3. The reason for the fall of the Exp1 index was the impact of the 2008 financial crisis in the U.S., but it was not prolonged and was only temporarily limited. In other words, it was an unpredictable decline. On the other hand, Exp2 and Exp3 are the results of reflecting the situation in which economic recession and interest rate fluctuations continue to affect the real estate market.

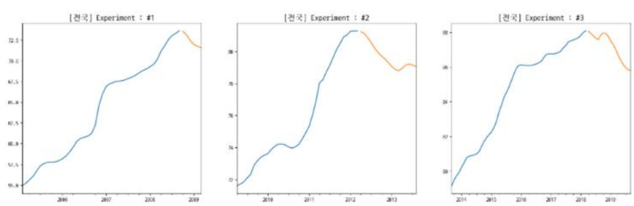

**Figure 3. Visualize national data sets**

## 4.2. Results of ARIMA

In order to perform ARIMA analysis, normality must be satisfied. In order to satisfy normality, it is generally converted into a normal time series through differences. Therefore, the national trading price index is tested by unit root for each experimental period to determine how much the difference will be. Table 2 shows the results of differentiating the national trading price index and performing a unit root test. It can be seen that the first difference in Exp1, the second difference in Exp2, and the second difference in Exp3 satisfy normality.

**Table 2. National ARIMA unit root test**

| Experiment | Difference | 1% | 5% | 10% | Augmented Dickey-Fuller (ADF) | P-Value(ADF) | Phillips Perron(tau) | P-Value(PP) |
|---|---|---|---|---|---|---|---|---|
| Exp1 | 1 | -3.600983 | -2.935135 | -2.605963 | 23.5928494 | 0.00367796 | -2.338997 | 0.15969883 |
| Exp2 | 2 | -3.646135 | -2.954127 | -2.615968 | 1.8130556 | 1.09E-11 | -8.119137 | 1.17E-12 |
| Exp3 | 3 | -3.571472 | -2.922629 | -2.599336 | -89.52074 | 2.39E-06 | -6.899508 | 1.29E-09 |

If the national trading price index is tuned (p, q) for each experiment, the results of Table 3 can be obtained. (p, q) is to select the value that AICc makes the smallest, and the optimal value was determined for each experiment. In Exp1, (2, 1, 0) was selected as the final model, in Exp2, (1, 2, 0) was selected as the final model, and in Exp3, (2, 2, 2) was selected as the final model.

**Table 3. National ARIMA parameters results**

| Experiment | pdq | AICc | Experiment | pdq | AICc | Experiment | pdq | AICc |
|---|---|---|---|---|---|---|---|---|
| Exp1 | (0, 1, 0) | 80.4649139 | Exp2 | (0, 2, 0) | -4.939921 | Exp3 | (0, 2, 0) | -110.8522 |
| Exp1 | (0, 1, 1) | 49.8537671 | Exp2 | (0, 2, 1) | -5.738012 | Exp3 | (0, 2, 1) | -109.0211 |
| Exp1 | (0, 1, 2) | 39.1048263 | Exp2 | (0, 2, 2) | -3.448121 | Exp3 | (0, 2, 2) | -108.7525 |
| Exp1 | (1, 1, 0) | 30.5673015 | Exp2 | (1, 2, 0) | -6.179635 | Exp3 | (1, 2, 0) | -109.0624 |
| Exp1 | (1, 1, 1) | 30.3174988 | Exp2 | (1, 2, 1) | -4.1112 | Exp3 | (1, 2, 1) | -106.4409 |
| Exp1 | (1, 1, 2) | 32.2241971 | Exp2 | (1, 2, 2) | -1.541076 | Exp3 | (1, 2, 2) | -107.828 |
| Exp1 | (2, 1, 0) | 30.0122543 | Exp2 | (2, 2, 0) | -3.976311 | Exp3 | (2, 2, 0) | -106.875 |
| Exp1 | (2, 1, 1) | 32.4467196 | Exp2 | (2, 2, 1) | -1.534967 | Exp3 | (2, 2, 1) | -105.2498 |
| Exp1 | (2, 1, 2) | 32.3770023 | Exp2 | (2, 2, 2) | -3.48178 | Exp3 | (2, 2, 2) | -114.2634 |

The national ARIMA performance indicators are shown in Table 4, and the national ARIMA model results are shown in Figure 4. As a result, it can be seen that ARIMA tends to predict higher than the actual value of the trading price index during the test period. In other words, it can be seen that the trend of falling prices is not reflected.

**Table 4. National ARIMA performance indices**

| Experiment | Exp1 | | Exp2 | | Exp3 | |
|---|---|---|---|---|---|---|
| Result | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| Training | 0.19389763 | 0.31551305 | 0.1393769 | 0.20684694 | 0.05758601 | 0.07435553 |
| Test | 1.78970295 | 2.00641706 | 1.74055341 | 1.94258753 | 1.82907008 | 2.2451975 |

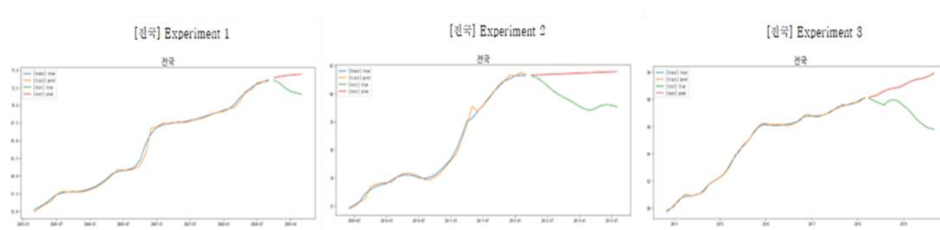


**Figure 4. National ARIMA model results**

The national ARIMA Accuracy visualization results are shown in Figure 5. The nationwide apartment sales price index varies greatly depending on economic conditions or events. Reflecting this point, the effective interval is set at 3%, 5%, and 10%, to check whether the corresponding predicted value comes into the actual value. It can be seen that Exp1, 2, and 3 are all predicting a continuous rise in the trading index. It was also confirmed that the volatility prediction ability was inferior when looking at the 3% error.

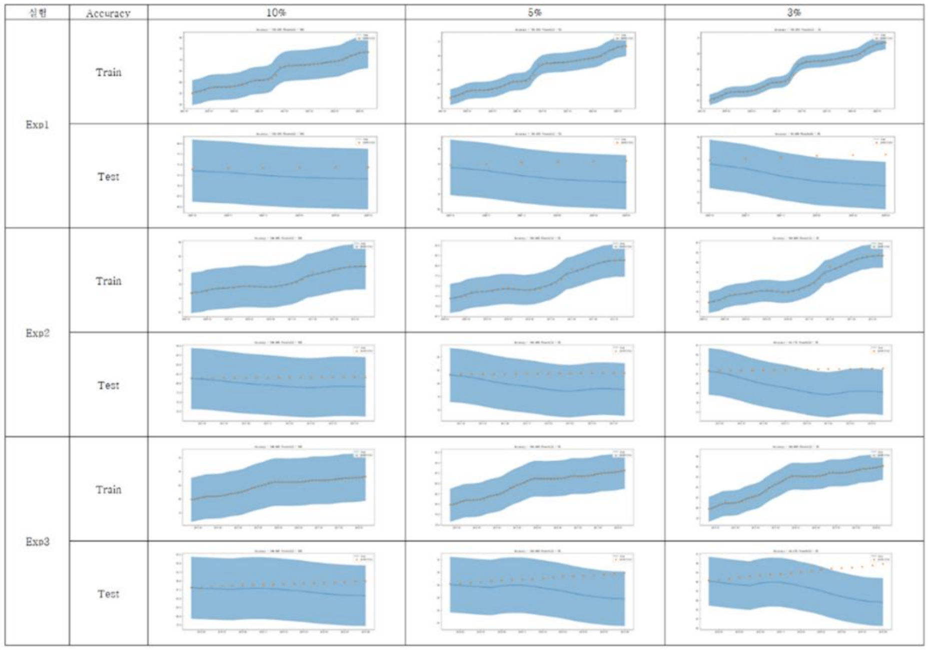


**Figure 5. National ARIMA accuracy visualization**

As a result of checking the Accuracy when predicted by ARIMA for each experiment, in Train, the accuracy of 100% is obtained by entering all sections, and in the test, the predictive power is low in some periods. It is predicted that the error will increase further if the test period increases (Table 5).

**Table 5. National ARIMA performance indices**

| Experiment | Accuracy | 10% | 5% | 3% |
|---|---|---|---|---|
| Exp1 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 50 |
| Exp2 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 64.7058824 |
| Exp3 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 66.6666667 |

### 4.3. Results of VAR

In the VAR model, the normality of the impact variable and the national apartment sales price index for each experiment must be satisfied. Therefore, the difference was carried out to summarize the values that satisfy normality. The national VAR unit root test is shown in Table 6. (A: the amount of money, B: the mining industry, C: prices, D: charter, E: corporate bonds, F: nationwide)

**Table 6. National VAR unit root test**

| Exp | Var. | Diff. | 1% | 5% | 10% | ADF | ADF | tau | PP |
|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 2 | -3.600983 | -2.935135 | -2.605963 | 600.88693 | 3.56E-12 | -8.042069 | 1.84E-12 |
| 1 | B | 1 | -3.596636 | -2.933297 | -2.604991 | 125.49048 | 1.21E-14 | -8.923258 | 1.03E-14 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | C | 1 | -3.596636 | -2.933297 | -2.604991 | 6.5406181 | 0.0001451 | -4.611351 | 0.0001231 |
| 1 | D | 2 | -3.632743 | -2.94851 | -2.613017 | -53.9767 | 0.0404084 | -4.849973 | 4.36E-05 |
| 1 | E | 1 | -3.596636 | -2.933297 | -2.604991 | -5.654716 | 0.0097679 | -3.405099 | 0.0107863 |
| 1 | F | 1 | -3.600983 | -2.935135 | -2.605963 | 23.592849 | 0.003678 | -3.043917 | 0.0309917 |
| 2 | A | 1 | -3.639224 | -2.95123 | -2.614447 | 479.72085 | 0.0055843 | -3.617908 | 0.0054278 |
| 2 | B | 0 | -3.65352 | -2.957219 | -2.617588 | 85.363148 | 0.0348209 | -2.852145 | 0.0512026 |
| 2 | C | 1 | -3.646135 | -2.954127 | -2.615968 | -0.178708 | 7.88E-10 | -4.348184 | 0.0003662 |
| 2 | D | 1 | -3.646135 | -2.954127 | -2.615968 | -14.50951 | 0.0184685 | -2.741132 | 0.067187 |
| 2 | E | 1 | -3.639224 | -2.95123 | -2.614447 | -20.96376 | 6.97E-05 | -4.771798 | 6.16E-05 |
| 2 | F | 2 | -3.646135 | -2.954127 | -2.615968 | 1.8130556 | 1.09E-11 | -7.68655 | 1.46E-11 |
| 3 | A | 1 | -3.560242 | -2.91785 | -2.596796 | 869.60226 | 1.90E-10 | -7.238186 | 1.91E-10 |
| 3 | B | 2 | -3.588573 | -2.929886 | -2.603185 | 154.61233 | 2.64E-05 | -24.30726 | 0 |
| 3 | C | 2 | -3.596636 | -2.933297 | -2.604991 | 10.728664 | 0.0010285 | -10.10095 | 1.06E-17 |
| 3 | D | 2 | -3.562879 | -2.918973 | -2.597393 | -103.9572 | 4.23E-12 | -8.014239 | 2.16E-12 |
| 3 | E | 1 | -3.560242 | -2.91785 | -2.596796 | -76.74196 | 2.71E-05 | -4.95318 | 2.74E-05 |
| 3 | F | 2 | -3.571472 | -2.922629 | -2.599336 | -89.52074 | 2.39E-06 | -6.56086 | 8.39E-09 |

Table 7 shows the results of determining the difference that satisfies normality for each variable and analyzing the influence between each variable through causal relationship analysis of the variables. As a result of the causal relationship analysis, it can be seen that the null hypothesis is rejected for the national apartment sales price index. Therefore, it can be seen that these variables statistically affect the nationwide apartment sales price index. Based on Lag8, Exp1 shows the order of VAR model arrangement in the order of mining industry, prices, corporate bonds, money supply, charter, and index. The order of arrangement of the Exp2 VAR models appears in the order of mining industry, corporate bonds, prices, money supply, charter, and index. The order of arrangement of the Exp3 VAR models appears in the order of charter, price, corporate bond, money supply, mining, and industry. (A: the amount of money, B: the mining industry, C: water prices, D: charter, E: corporate bonds, F: nationwide)

**Table 7. National granger causality analysis**

| Experiment | | | Exp1 | | | Exp2 | | | Exp3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Null | | | P-Value | | | P-Value | | | P-Value | | |
| | | | lag2 | lag4 | lag8 | lag2 | lag4 | lag8 | lag2 | lag4 | lag8 |
| A | -X→ | B | 0.2583 | 0.4346 | 0.7929 | 0.6943 | 0.9442 | 0.0871 | 0.0101 | 0.0289 | 0.0058 |
| A | -X→ | C | 0.0233 | 0.1538 | 0.0915 | 0.0047 | 0.0891 | 0.0994 | 0.4299 | 0.8118 | 0.0176 |
| A | -X→ | D | 0.1216 | 0.2448 | 0.0132 | 0.6978 | 0.3122 | 0 | 0.5486 | 0.5326 | 0.3769 |
| A | -X→ | E | 0.1993 | 0.4085 | 0.0381 | 0.6732 | 0.2465 | 0.0001 | 0.2144 | 0.0114 | 0.0219 |
| A | -X→ | F | 0.2709 | 0.1343 | 0.0917 | 0.7379 | 0.217 | 0.2205 | 0.7162 | 0.8635 | 0.4983 |
| B | -X→ | A | 0.3045 | 0.5933 | 0.1285 | 0.0705 | 0.0592 | 0.1334 | 0.6069 | 0.0149 | 0.1177 |
| B | -X→ | C | 0.3657 | 0.3375 | 0.3006 | 0.2221 | 0.4952 | 0.0406 | 0.9669 | 0.7265 | 0.5434 |
| B | -X→ | D | 0.3236 | 0.534 | 0.045 | 0.4818 | 0.4078 | 0.077 | 0.1955 | 0.5877 | 0.0187 |
| B | -X→ | E | 0.518 | 0.4797 | 0.1294 | 0.0643 | 0.3786 | 0.2396 | 0.6222 | 0.2254 | 0.6114 |
| B | -X→ | F | 0.8131 | 0 | 0 | 0.042 | 0.0007 | 0 | 0.1235 | 0.4497 | 0.0875 |
| C | -X→ | A | 0.5728 | 0.2433 | 0.0177 | 0.0375 | 0.001 | 0 | 0.0358 | 0.0779 | 0.1759 |
| C | -X→ | B | 0.0061 | 0.0532 | 0 | 0.7426 | 0.959 | 0.0015 | 0.7266 | 0.8863 | 0.1709 |
| C | -X→ | D | 0.1504 | 0.5329 | 0.0195 | 0 | 0 | 0 | 0.4979 | 0.3426 | 0.6279 |
| C | -X→ | E | 0.0195 | 0.0354 | 0.0829 | 0.6097 | 0.7096 | 0.0012 | 0.8721 | 0.8221 | 0.0336 |
| C | -X→ | F | 0.3497 | 0.2686 | 0.0273 | 0.0719 | 0.0017 | 0.0003 | 0.8263 | 0.3725 | 0.4287 |
| D | -X→ | A | 0.6599 | 0.982 | 0.1619 | 0.0486 | 0.032 | 0.0023 | 0.882 | 0.8502 | 0.9904 |
| D | -X→ | B | 0.0996 | 0.0036 | 0 | 0.5842 | 0.4887 | 0.0029 | 0.7718 | 0.4192 | 0.7065 |
| D | -X→ | C | 0.0033 | 0.0685 | 0.0044 | 0.0764 | 0.1798 | 0.002 | 0.8272 | 0.8861 | 0.5598 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| D | -X→ | E | 0.2436 | 0.0786 | 0.1681 | 0.611 | 0.8813 | 0.19 | 0.0103 | 0.0589 | 0.0438 |
| D | -X→ | F | 0.0029 | 0.012 | 0 | 0.0127 | 0.0469 | 0 | 0.2635 | 0.0959 | 0.4725 |
| E | -X→ | A | 0.5429 | 0.0639 | 0.0125 | 0.7277 | 0.508 | 0.0296 | 0 | 0 | 0.0008 |
| E | -X→ | B | 0.5271 | 0.2891 | 0.9572 | 0.7648 | 0.115 | 0.0079 | 0.8649 | 0.0265 | 0.0113 |
| E | -X→ | C | 0.7696 | 0.7225 | 0.5538 | 0.1722 | 0.0212 | 0 | 0.8374 | 0.9342 | 0.0002 |
| E | -X→ | D | 0.116 | 0.417 | 0.1913 | 0.0604 | 0.2723 | 0 | 0.7892 | 0.2015 | 0 |
| E | -X→ | F | 0.1379 | 0.1514 | 0.3669 | 0.1744 | 0.2717 | 0.0683 | 0.4933 | 0.6793 | 0.4695 |
| F | -X→ | A | 0.7024 | 0.6871 | 0.341 | 0.3294 | 0.3022 | 0.0001 | 0.713 | 0.4091 | 0.1224 |
| F | -X→ | B | 0.001 | 0.0012 | 0.0248 | 0.7072 | 0.9633 | 0.0002 | 0.8039 | 0.3622 | 0.6612 |
| F | -X→ | C | 0.8601 | 0.2832 | 0.5456 | 0.5286 | 0.0098 | 0 | 0.2476 | 0.2046 | 0.2516 |
| F | -X→ | D | 0.1744 | 0.2347 | 0.0204 | 0.8755 | 0.2721 | 0.0179 | 0.1216 | 0.0586 | 0.0248 |
| F | -X→ | E | 0.2176 | 0.85 | 0.8117 | 0.5676 | 0.9378 | 0.5502 | 0.0784 | 0.2313 | 0.0689 |

Table 8 shows the national AIC, SC, and HQIC results. In order to find the optimal order for each experiment, the candidate group was selected by changing the order of the lag order, and the lag order with the lowest AIC was selected. Lag Order 6 was selected in Exp1, Lag Order 2 was selected in Exp2, and Lag Order 6 was selected in Exp3. And Table 9 shows the national VAR performance indicators.

**Table 8. National AIC, SC, and HQIC results**

| Experiment | Lag order | AIC | SC | HQIC |
|---|---|---|---|---|
| Exp1 | 0 | 6.6129168 | 6.8611553 | 6.703906 |
| | 1 | 5.7119804 | 7.4673469 | 6.3511881 |
| | 2 | 5.5501107 | 8.8434256 | 6.7408694 |
| | 3 | 5.0268351 | 9.8895537 | 6.7715367 |
| | 4 | 3.5505509 | 10.014707 | 5.8504482 |
| | 5 | -6.141163 | 1.9569645 | -3.286194 |
| | 6 | -301.939 | -292.174 | -298.5307 |
| Exp2 | 0 | 6.6902348 | 6.9595926 | 6.7820935 |
| | 1 | 6.5807788 | 8.4854247 | 7.2216342 |
| | 2 | 5.8742758 | 9.447007 | 7.0585351 |
| | 3 | 5.1249735 | 10.398346 | 6.8439614 |
| Exp3 | 0 | 3.4812891 | 3.704341 | 3.5670641 |
| | 1 | 2.8663614 | 4.4423659 | 3.4705643 |
| | 2 | 2.4852469 | 5.4398038 | 3.6142705 |
| | 3 | 1.4935215 | 5.8529339 | 3.1536106 |
| | 4 | 1.7906846 | 7.5819713 | 3.9878907 |
| | 5 | 0.5536851 | 7.8045891 | 3.2938123 |
| | 6 | -3.184989 | 5.5540058 | 0.1035533 |

**Table 9. National VAR performance indices**

| Experiment | Exp1 | | Exp2 | | Exp3 | |
|---|---|---|---|---|---|---|
| Result | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| Training | 1.2308E-12 | 1.5718E-12 | 0.06916153 | 0.09566924 | 0.03151975 | 0.0394792 |
| Test | 0.8485175 | 1.0455586 | 0.20444991 | 0.25626971 | 0.13680306 | 0.16759344 |

The results of the national VAR model were shown in Figure 6. Through this, it can be seen that VAR has accurately derived trend analysis compared to ARIMA. However, in the process of being affected by variables at the same time, the trend of the trading price index falling was not consistently predicted. The sharp change in the Exp1 forecast was affected by the small number of data and the increased volatility of variables such as corporate bond yields and mining industry indices in the 2008 financial crisis.

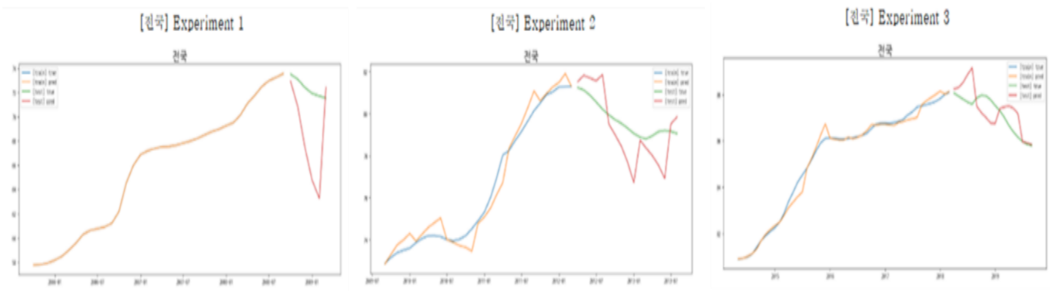


**Figure 6. National VAR model results**

The national VAR Accuracy visualization results are shown in Figure 7, and it was confirmed that the Exp1 predictive power was low. It can be seen that the predictive power is lower than that of ARIMA (3%, 50). However, Exp2 and 3 showed higher predictive power than ARIMA.

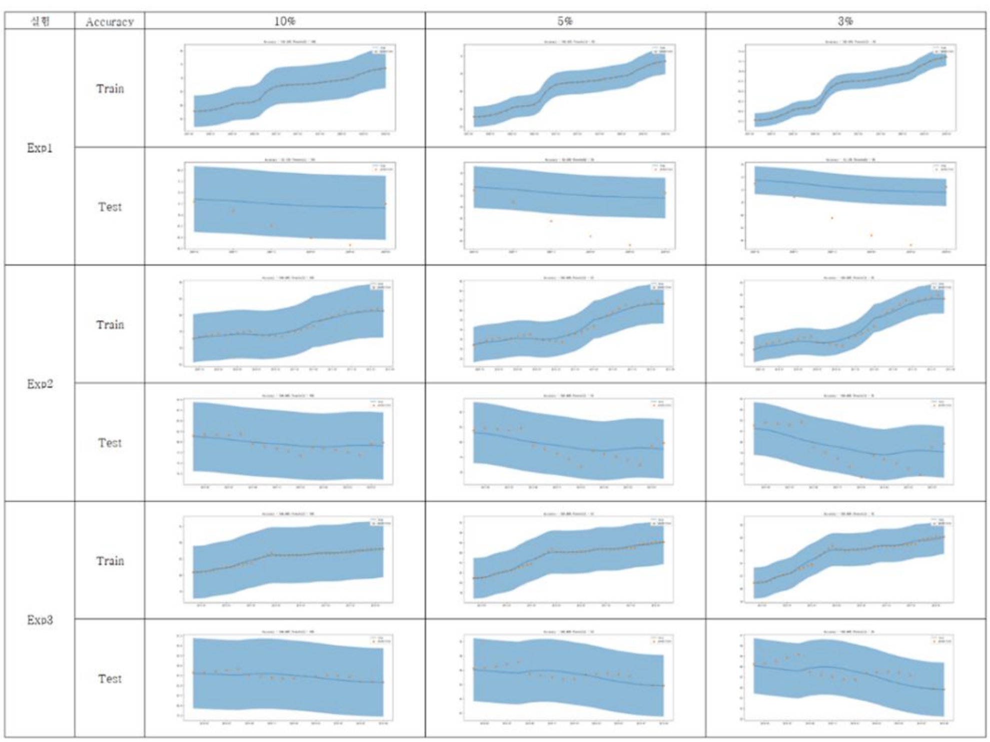


**Figure 7. National VAR accuracy visualization**

As a result of the national VAR Accuracy performance indicators, it can be seen that the Exp1 predictive power is poor through Table 10. The impact is that the number of data is small, and the volatility of variables has increased due to the 2008 financial crisis. In addition, the economic situation at that time seems to have had an impact on the external shock, not the domestic situation.

**Table 10. National VAR performance indices**

| Experiment | Accuracy | 10% | 5% | 3% |
|---|---|---|---|---|
| Exp1 | Training | 100 | 100 | 100 |
| | Test | 83.3333333 | 50 | 33.3333333 |
| Exp2 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 100 |
| Exp3 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 100 |

### 4.4. Results of LSTM

When looking at the LSTM performance indicators, it was confirmed that the national trading price index has better predictive power than ARIMA and VAR (Table 11). Considering the predictive power, it was possible to predict the decline in the trading price index in the case of Exp1 and 3. In the case of Exp2, the downward trend was not predicted, but it is predicted that the range of the rise will decrease.

**Table 11. National LSTM performance indices**

| Experiment | Exp1 | | Exp2 | | Exp3 | |
|---|---|---|---|---|---|---|
| Result | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| Training | 0.00435058 | 0.00564518 | 0.00403324 | 0.0055949 | 0.00428252 | 0.00557655 |
| Test | 0.49665188 | 0.63971511 | 2.3860618 | 2.65691774 | 0.71933329 | 0.92489999 |

As a result of the national LSTM model in Figure 8, the loss of the train is falling normally, indicating that the train has been carried out normally. In the case of Exp3, the trading price index did not consistently fall from April 2014 to August 2013, but rather fell after a slight rebound, and this volatility was also reflected in the LSTM forecast.

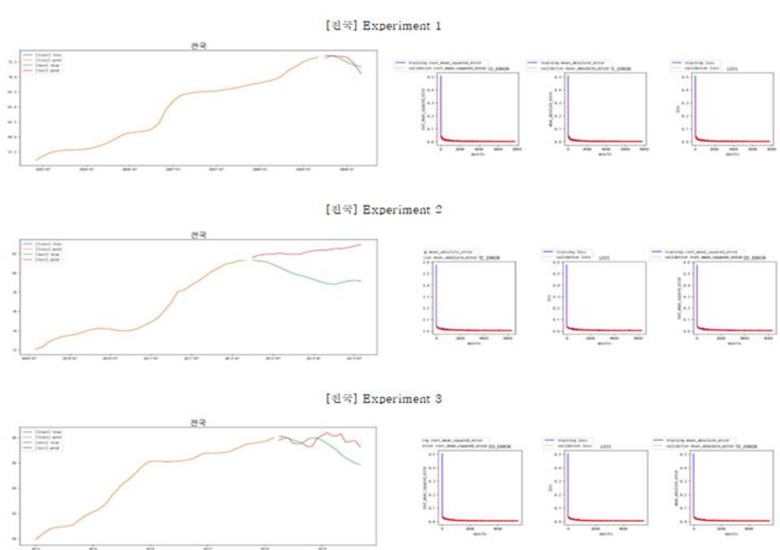


**Figure 8. National LSTM model results**

The national LSTM Accuracy visualization results are shown in Figure 9 and Table 12. At a time when the volatility corresponding to the test is high, LSTM similarly predicts regulators in Exp1, and Exp2 tends to predict higher than the actual value of the trading price index. For Exp2, it was not predicted within the 3% error range.

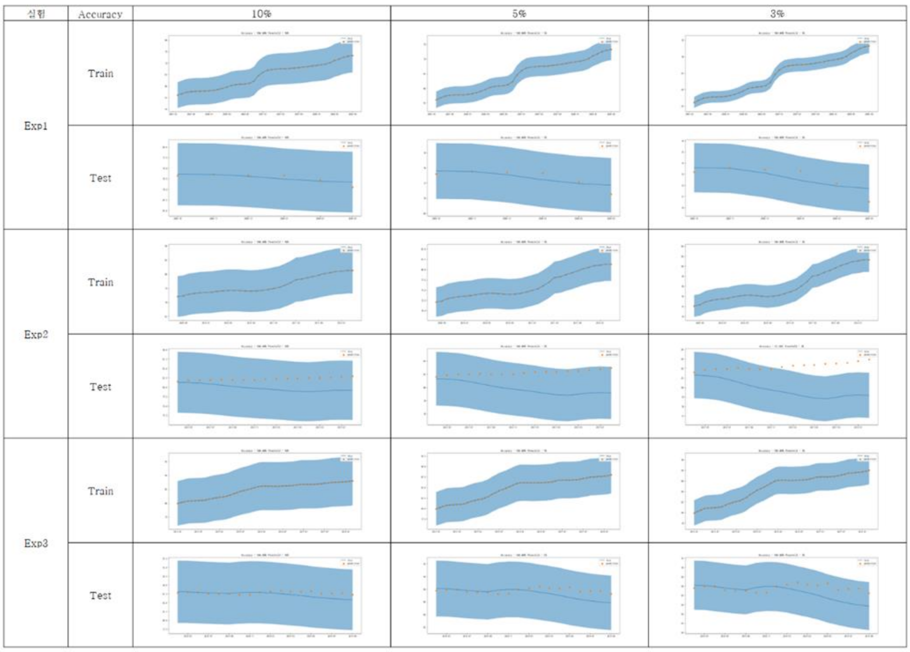


**Figure 9. National LSTM accuracy visualization**

**Table 12. National LSTM accuracy performance indices**

| Experiment | Accuracy | 10% | 5% | 3% |
|---|---|---|---|---|
| Exp1 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 100 |
| Exp2 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 47.0588235 |
| Exp3 | Training | 100 | 100 | 100 |
| | Test | 100 | 100 | 100 |

## 5. Conclusions

The period in which the index showed a downward trend was Exp1(6 months), Exp2(17 months), and Exp3(18 months), and the period and width of volatility were different for each experiment. Exp1 was the result of the economic downturn and interest rates continuing to affect the real estate market through the impact of the financial crisis from the United States. In the case of ARIMA, all three cases tended to be higher than the actual value, and as a result, the downward trend was not predicted. VAR was able to predict the downward trend, but the prediction fluctuated greatly as several variables affected it at the same time. In particular, it was difficult to predict the sudden shock (financial crisis from the United States). Therefore, LSTM predicted the downward trend more accurately than ARIAM and VAR, and it was found that the trend of fluctuations can be predicted. Experimenting with periods of increased volatility confirmed that LSTM using artificial intelligence is generally accurate and especially predicts fluctuation trends in indices by following them faster than other methodologies. However, the performance of LSTM was not always excellent. ARIAM's predictive power was high when the index change moved in a constant trend rather than an external shock. However, it was confirmed that ARIMA, a traditional technique, is less predictive in rapidly changing economic conditions, and even in the case of LSTM, the trend can be confirmed. In other words, it was confirmed that the LSTM was excellent in predicting the start of the fluctuation period. As a result, in the case of prediction through LSTM, it was confirmed that it was effective to use it in parallel with traditional time series techniques such as ARIMA.

## Acknowledgement

This paper was supported by Wonkwang University in 2022.

## References

[1] M. A. Boden, "Artificial intelligence", Elsevier, 2023.
[2] Y. Mintz, and R. Brodie, "Introduction to artificial intelligence in medicine", Minimally Invasive Therapy & Allied Technologies, Vol. 28, No. 2, pp. 73-81, 2019.
DOI: https://doi.org/10.1080/13645706.2019.1575882
[3] S.H. Kim, J.K. Choi, J.S. Kim, A.R. Jang, J.H. Lee, K.J. Cha, and S.W. Lee, "Animal Infectious Diseases Prevention through Bigdata and Deep Learning", Journal of Intelligence and Information Systems, Vol. 24. No. 4. pp. 137-154, 2018.
DOI: https://doi.org/10.13088/jiis.2018.24.4.137
[4] C. H. Lee, K. Park, and S. Lee, "Development of IoT Healthcare Platform Model for the Elderly using Bigdata and Artificial Intelligence", International Journal of Membrane Science and Technology, Vol. 10, No. 1, pp. 108-113, 2023.
https://doi.org/10.15379/ijmst.v10i1.1435
[5] D. Fasel, and A, Meier, Big data, Springer Vieweg, 2014
[6] S. Lee and J. Kim, "Data Design Strategy for Data Governance Applied to Customer Relationship Management", International Journal of Advanced Culture Technology, Vol. 11, No. 3, pp. 338-345, 2023.
DOI: https://doi.org/10.17703/IJACT.2023.11.3.338
[7] S.H. Kim, S. Chang, and S.W. Lee, "Consumer Trend Platform Development for Combination Analysis of Structured and Unstructured Bigdata", Journal of Digital Convergence, Vol. 15. No. 6. pp. 133-143, 2017.
DOI: https://doi.org/10.14400/JDC.2017.15.6.133
[8] S. Sagiroglu, and D. Sinanc, "Big data: A review", The 2013 international conference on collaboration technologies and systems (CTS), IEEE, pp. 42-47. 2013, May.
DOI: http://doi.org/10.1109/CTS.2013.6567202
[9] R.H. Shumway, et al. "ARIMA models", Time Series Analysis and its Applications: with R Examples, pp. 75-163, 2017.
DOI: https://doi.org/10.1007/978-3-319-52452-8_3
[10] H. Lütkepohl, "Vector autoregressive models", Handbook of Research Methods and Applications in Empirical Macroeconomics, Vol. 30, 2013.
[11] R.C. Staudemeyer and E.R. Morris, "Understanding LSTM - a Tutorial into Long Short-Term Memory Recurrent Neural Networks", arXiv Preprint arXiv:1909.09586, 2019.
DOI: https://doi.org/10.48550/arXiv.1909.09586