IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 27, 2021, accepted April 29, 2021, date of publication May 6, 2021, date of current version June 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3077962

A Machine Learning-Based Early Warning System for the Housing and Stock Markets

DAEHYEON PARK AND DOOJIN RYU
College of Economics, Sungkyunkwan University, Seoul 03063, South Korea

Corresponding author: Doojin Ryu (sharpjin@skku.edu)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government, Ministry of Science and ICT (MSIT), under Grant 2019R1G1A1100196.

ABSTRACT This study analyzes the relationship between the housing and stock markets, focusing on housing market bubbles. Stock market dynamics generally have a more significant impact on housing price movements than housing market dynamics have on stock dynamics. However, if housing market information is provided as a signal, housing price movements can predict stock market volatility. Accordingly, we build a machine learning-based early warning system (EWS) for the housing market using a long short-term memory (LSTM) neural network. Applying the generalized supremum augmented Dickey-Fuller test to extract the bubble signal in the housing market, we find that the signal simultaneously detects future changes in the housing market prices and future stock market volatility, and our EWS effectively detects the bubble signal. We confirm that the LSTM approach performs better than other benchmark models, the random forest and support vector machine models.

INDEX TERMS Early warning system, housing market bubble, long short-term memory, machine learning, stock market volatility.

I. INTRODUCTION

Housing market bubbles are widely recognized but present an intractable risk [1], [2]. When speculation becomes rampant based on investors' expectations and sentiment, a bubble forms in the housing market because the housing supply is inelastic. Investors tend to ignore that a rise in housing prices may be a bubble rather than an increase in housing's intrinsic value. In the Korean real estate market, housing is historically preferred to other assets, as it offers higher profitability compared to deposits and bears lower risk compared to stocks [3], [4]. Because of the downward rigidity of housing prices, housing is considered a stable investment [5]. Furthermore, the housing market supply is inelastic because it usually takes more than several years to construct the houses. Thus, the excess demand in the housing market is likely to cause prices to continuously rise, likely to result in housing market bubbles.

A housing market bubble indicates that the market is unstable owing to abnormally high housing prices. The collapse of a housing market bubble impacts the country's economy, including its financial market [6], [7]. The global financial crisis triggered by the subprime mortgage crisis shows that housing bubble bursting can spread throughout the financial market, compounding its damage. Thus, housing market bubbles must be accurately predicted and effectively managed. The Korean government's real estate policy either accompanies or lags behind economic fluctuations, which may be the reason for its failure [8], [9]. Time-series models based on the Box-Jenkins methodology, such as the autoregressive moving average and autoregressive integrated moving average models, are the primary means of predicting time-series data. The data stability and stationarity must be ensured for these models to obtain significant and meaningful estimates. However, it is difficult to directly apply the classical time-series models to the housing market because housing prices are unstable and can be affected by various external variables, including real estate policies [10].

One way to overcome these limitations and effectively predict housing market bubbles is an early warning system (EWS) [11]. We propose a new, effective EWS using a long short-term memory (LSTM) neural network, a type

The associate editor coordinating the review of this manuscript and approving it for publication was Michael Lyu.

of machine learning methodology. We define the signal for a housing market bubble following Phillips *et al.* [12] and predict the signal using the LSTM neural network. By demonstrating the link between the EWS signal and financial market volatility, we show that our system can serve as a comprehensive EWS for the housing and stock markets.

The remainder of this paper is organized as follows. Section II presents the theoretical background for this study, including details on EWSs in the housing and financial markets. Section III constructs a housing market bubble signal and evaluates whether the signal significantly predicts financial market volatility. Section IV develops the housing market EWS using the LSTM neural network. Finally, Section V summarizes and concludes the study.

II. BACKGROUND AND LITERATURE REVIEW

An EWS is designed to detect signals of a future crisis in specific markets in order to proactively respond to the crisis. The need to preemptively respond to crises has increased since the Asian currency crisis in the 1990s, and so does the research on EWSs, accordingly. Various studies suggest approaches to construct effective EWSs. The primary methods used in the existing literature are the signal approach suggested by Kaminsky *et al.* [13] and the probit model presented by Frankel and Rose [14]. The signal approach proposes a model that constructs indicators using all available variables and predicts the occurrence of a crisis based on whether these indicators generate signals. The probit model seeks to determine the conditional probability of a crisis based on the relationships among a given set of variables and the likelihood that a crisis will occur.

Current research on EWSs focuses more on financial or banking crises [15]. Aldasoro *et al.* [16], for example, construct early warning indicators using a combination of debt variables and housing prices to signal banking crises. However, the literature includes only a few early warning models that focus on the housing market itself. Dreger and Kholodilin [17] use several macroeconomic and financial variables to construct an EWS for speculative housing market bubbles. Ferrari *et al.* [18] test various early warning signals for real estate-related banking crises.

Our study establishes a practical EWS for the Korean housing market using an LSTM neural network. We develop a dynamic EWS by integrating a crisis classifier and the LSTM neural network. The LSTM neural network is proven to be a state-of-the-art mechanism in the general financial forecasting field [19]–[21], including the predictions of future stock price [22]–[24] and volatility [25], [26]. Sezer *et al.* [27] systematically review the literature on financial forecasting with deep learning during the period 2005–2019 and conclude that models based on recurrent neural networks (RNNs), such as LSTM and gated recurrent unit networks, are the most commonly accepted models because they dominate in price and trend predictions and are well adapted to all sorts of forecasting problems. To the best of our knowledge, this study is the first to incorporate an LSTM neural network into an EWS.

Existing studies that design EWSs focus only on their predictive power for the housing market, and no previous study has examined whether early warning signals can also indicate volatility in the financial market. Given that a financial crisis triggered by a housing market crisis has a ripple effect, it is necessary to design an EWS that can simultaneously detect risk in both the housing and stock markets. Although many studies show that the housing and financial markets are significantly correlated, the findings regarding the direction of causality are inconsistent. Empirically investigating Taiwan's stock market, Chen [28] argues that the stock market has a one-way causal effect on the real estate market. Okunev *et al.* [29] analyze the dynamic relationship between the U.S. property market and the S&P 500 stock market and suggest an opposite conclusion. However, using bubble signals, which are refined information, rather than housing prices, can yield different results. Thus, this study analyzes whether stock market volatility is asymmetric to an EWS's signals, which would suggest that the housing market bubble can significantly explain stock market characteristics.

III. HOUSING MARKET BUBBLE AND STOCK MARKET VOLATILITY

A. HOUSING MARKET BUBBLE

This section defines the housing market bubble signal for building an EWS, using the housing market pressure index (*HMPI*). *HMPI* consists of housing price indicators because the housing market instability is observed through the housing price. This study's housing price indicators include the Korean housing sales price (KSI_t) and *Chonsei* price (KCI_t) indexes and the housing sales price (GSI_t) and *Chonsei* price (GCI_t) indexes of *Gangnam*, Seoul.[1] *HMPI* may be configured in various ways using these variables; in this study, *HMPI* is calculated as the maximum of the four variables. This method is advantageous as *HMPI* may not be affected even if a single indicator is excessively low. Accordingly, *HMPI* in this study is defined as follows.

$$HMPI_t = MAX\{KSI_t, KCI_t, GSI_t, GCI_t\} \tag{1}$$

An *HMPI* value above a certain level signals instability in the housing market. This signal can be interpreted as a sign of a housing bubble because it is highly likely to be followed by a rapid fall in housing prices. Some studies define the occurrence of a crisis using the average and standard deviation

[1] *Chonsei* is a unique housing rental system in Korea. A tenant taking a *Chonsei* contract normally pays a deposit lower than the housing price in advance and gets the deposit back at maturity [30]. Unlike typical housing rental systems in other countries, the tenant does not need to pay periodic rental payments. In the Korean housing market, there are *Chonsei* contracts as many as sales contracts, so *Chonsei* contracts are often included when analyzing the housing market. *Gangnam* is a region in Seoul, Korea, famous for Psy's K-pop song "*Gangnam Style*." It is also known for expensive real estate prices with convenient transportation, shopping and business centers, and elite schools and academies. Importantly, housing prices in the *Gangnam* region have triggered bubbles and co-movements in other regional housing markets [31]. Many studies show that the rise in Korean housing prices is triggered by the *Gangnam* housing market [32], implying the market's considerable influence in Korea's real estate market. Accordingly, this study uses variables from the *Gangnam* district to construct *HMPI*.

over an entire period [33]. However, this metric only provides a significant signal if the market pressure index is a stable time series with no trend. In case the market pressure index instead has a constant trend, then the market's upward and downward phases will have different characteristics. When the mean and standard deviation of the data for an entire period are employed, the threshold for determining instability is extremely low (high) in the ascending (descending) phase. Thus, it is difficult to distinguish the price increases caused by market booms from those caused by bubbles. Therefore, building an effective EWS requires constructing a short-term comparative sample so that the system can capture the signals of crises before they occur.

Following Hagemann and Wohlmann [10], we apply the generalized supremum augmented Dickey-Fuller (GSADF) test, which is based on the augmented Dickey-Fuller (ADF) regression model. The ADF model tests for a unit root process within the time-series data. The GSADF test is the recursive ADF test that examines whether the time-series data follow a random walk (i.e., a unit root) or an explosive process, which indicates a price bubble. The recursive ADF test regression is estimated on rolling-window subsamples, where the starting point, r_1, varies from 0 to $r_2 - r_{\mathrm{w}}$; the endpoint, r_2, varies from r_w to 1. r_w is the minimum subsample size. We set r_w equal to 20% of the full sample size. Thus, GSADF statistic is defined as the supremum of ADF statistics from the starting point, r_1, to the endpoint, r_2, as follows.

$$GSADF\ (r_w) = \sup_{\substack{r_2 \in [r_w, 1] \\ r_1 \in [0, r_2 - r_w]}} \{ADF_{r_1}^{r_2}\} \tag{2}$$

The bubble period is estimated using the backward supremum augmented Dickey-Fuller (BSADF) statistic. We obtain a sequence of ADF statistics for every endpoint r_2, and the supreme value of each sequence yields the BSADF statistic.

$$BSADF_{r_2}\ (r_w) = \sup_{r_1 \in [0, r_2 - r_w]} \{ADF_{r_1}^{r_2}\} \tag{3}$$

The origination date, $\hat{r_e}$, of the price bubble is defined as the first observation date, whose BSADF statistic exceeds the critical value cv at a given significance level α. The termination date, $\hat{r_f}$, is the first observation date after the origination date, whose BSADF statistic falls below cv.

$$\hat{r_e} = \inf_{r_2 \in [r_w, 1]} \{r_2 : BSADF_{r_2}\ (r_w) > cv_{r_2}^{\alpha}\} \tag{4}$$

$$\hat{r_f} = \inf_{r_2 \in [\hat{r}_e + \log(T)/T, 1]} \{r_2 : BSADF_{r_2}\ (r_w) > cv_{r_2}^{\alpha}\} \tag{5}$$

The housing market bubble signal (*HBS*) is a dummy variable that equals 1 for the period from the origination date to the termination date and 0 otherwise. Figure 1 shows the bubble signal in the housing market.

We use the regression model given by Equation (6) to analyze whether *HBS* significantly affects future changes in *HMPI*.

$$HMPI_{t+k} = \alpha_k + \beta_k HBS_t + \epsilon_k, \quad \epsilon \sim N\left(0, \sigma_k^2\right) \tag{6}$$

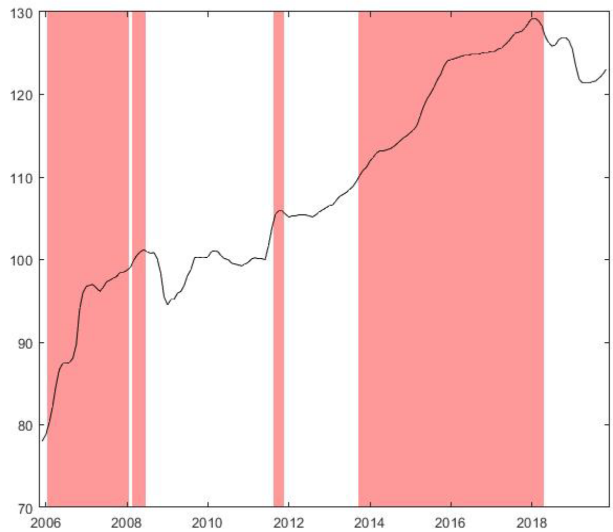

FIGURE 1. HMPI and HBS. Crisis periods determined by the HBS are highlighted in red.

TABLE 1. Effect of HBS on HMPI.

	(A) *HMPI* (β_k)		(B) *dHMPI* (b_k)	
k=1	5.8291***	(0.0027)	0.3218***	(0.0048)
k=2	6.2724***	(0.0011)	0.2349**	(0.0415)
k=3	6.6664***	(0.0004)	0.1820	(0.1143)
k=4	7.0190***	(0.0002)	0.1374	(0.2304)
k=5	7.3709***	(0.0001)	0.1334	(0.2322)
k=6	7.6689***	(0.0000)	0.0759	(0.4924)
k=7	7.9370***	(0.0000)	0.0425	(0.7024)
k=8	8.2142***	(0.0000)	0.0482	(0.6663)
k=9	8.4453***	(0.0000)	-0.0015	(0.9895)
k=10	8.5719***	(0.0000)	-0.1093	(0.3342)
k=11	8.5929***	(0.0000)	-0.2177*	(0.0521)
k=12	8.6191***	(0.0000)	-0.2145**	(0.0299)

Note: This table shows the effect of *HBS* on *HMPI* and *dHMPI*. *k* is the time lag. β_k and b_k are the coefficient estimates. Numbers in parentheses are *p*-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Equation (6) shows the effect of *HBS* on *HMPI* after *k* months. Table 1 (A) summarizes the estimated coefficient of β_k and the *p*-value for each time difference $k = 1, \ldots, 12$. In Table 1 (A), all of β_k estimates are significantly positive, and the β_k estimate increases as the time difference increases. Thus, after the signal (i.e., $HBS = 1$) appears, *HMPI* rises rapidly. That is, *HBS* can detect bubbles in *HMPI*.

To robustly confirm this result, we estimate Equation (7), where *dHMPI* is the first-differenced value of *HMPI*.

$$dHMPI_{t+k} = a_k + b_k HBS_t + e_k, \quad e_k \sim N\left(0, \sigma_k^2\right) \tag{7}$$

As shown in Table 1 (B), the b_k estimates are significantly positive when $k = 1, 2$. It turns negative when $k = 9$ and significantly negative when $k = 11, 12$. In summary, we can conclude that *HMPI* increases rapidly after the origination date ($\hat{r_e}$) and then the growth rate decreases after about a year. In other words, *HMPI* quickly approaches close to a maximum after the origination date, and such movements pose the risk of an imminent fall. *HBS* in this study, therefore, is an appropriate indicator of instability in the Korean housing market.

B. HOUSING BUBBLE SIGNAL AND STOCK MARKET VOLATILITY

As discussed in Section II, many studies find that housing and stock markets are significantly correlated. In general, prior studies argue that the stock market's impact on the housing market is greater than that of the housing market on the stock market.

In contrast to these studies, this section shows that risk signals in the housing market significantly impact stock market volatility. If *HBS* asymmetrically affects stock market volatility, then our EWS for the housing market can also detect some risk signals in the stock market. Because significant damage follows when a housing market crisis causes a financial crisis, as in the case of the global financial crisis, our EWS can be more helpful in responding to an economic crisis than other existing EWSs are.

Because stock market volatility is not an observed variable, we need to develop a separate volatility measure. We define the monthly realized volatility as follows.

$$rtn_d = ln\,(price_d) - ln(price_{d-1}) \quad (8)$$

$$RV_m = \sqrt{\frac{1}{T}\sum\nolimits_{d=1}^{T} rtn_d^2} \quad (9)$$

Equation (8) is the formula for the daily yield. The daily rate of return (rtn_d) is calculated as the daily logarithm return of the Korea Composite Stock Price Index (KOSPI) 200 index ($price_d$) from November 2004 to December 2019. Equation (9) calculates the monthly realized volatility (RV_m) as the square root of the mean of the squared rtn_d over the month.

We use the bias test proposed by Engle and Ng [34] to analyze the impact of instability signals in the housing market on the calculated monthly volatility. We conduct a sign bias test, a negative size bias test, a positive size bias test, and a joint bias test, as described by Equations (10) through (13). These tests are employed to confirm that the volatility is asymmetric and reacts more sensitively to negative shocks.

$$sign\ bias\ test: \epsilon_t^2 = \alpha + \beta \cdot S_{t-1}^- + e_t \quad (10)$$

$$negative\ size\ bias\ test: \epsilon_t^2 = \alpha + \beta \cdot S_{t-1}^- \cdot \epsilon_{t-1} + e_t \quad (11)$$

$$positive\ size\ bias\ test: \epsilon_t^2 = \alpha + \beta \cdot S_{t-1}^+ \cdot \epsilon_{t-1} + e_t \quad (12)$$

$$joint\ bias\ test: \epsilon_t^2 = \alpha + \beta_1 \cdot S_{t-1}^- + \beta_2 \cdot S_{t-1}^- \cdot \epsilon_{t-1} + \beta_3 \cdot S_{t-1}^+ \cdot \epsilon_{t-1} + e_t \quad (13)$$

ϵ_t is the residual, and S_t^- (S_t^+) is a dummy variable equal to one when $\epsilon_t < 0$ ($\epsilon_t > 0$). For Equations (10), (11), and (12), we perform a t-test on the coefficient β and conclude that the bias exists if the coefficient is statistically significant. For Equation (13), we perform an F-test and conclude that bias exists if the test result is statistically significant. We conduct four bias tests so that we can comprehensively interpret the results. If volatility reacts more sensitively to negative shocks, then the sign bias (Equation (10)), negative size bias (Equation (11)), and joint bias tests (Equation (13)) will be significant, but the positive size bias test (Equation (12)) will not be significant. We test the residuals for stock market volatility bias for signals in the housing market by modifying the above test to replace the dummy variables with the variable *HBS*.

TABLE 2. Bias test.

	Test 1	Test 2	Test 3	Test 4
***k* = 1**※	-2.763***	0.458	8.352***	65.4***
***k* = 2**※	-2.827***	0.023	6.134***	25.8***
***k* = 3**※	-2.712***	-0.289	5.901***	20.9***
***k* = 4**※	-2.035**	0.054	4.051***	10.4***
***k* = 5**※	-1.833*	0.170	3.091***	6.78***

Note: This table shows the results of the bias tests for stock market volatility in response to housing market signals for k=1, …, 5. k is the time lag, and the numbers in the cells are test statistics. Columns *Test 1*, *Test 2*, and *Test 3* report the t-statistics for the coefficients in Equations (14)–(16), respectively. Column *Test 4* reports the F-statistics for the coefficients in Equation (17). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. ※ indicates that the signal of instability in the housing market causes asymmetric volatility in the stock market.

$$Test\ 1: RV_t^2 = \alpha + \beta \cdot HBS_{t-k} + e_t \quad (14)$$

$$Test\ 2: RV_t^2 = \alpha + \beta \cdot HBS_{t-k} \cdot RV_{t-k} + e_t \quad (15)$$

$$Test\ 3: RV_t^2 = \alpha + \beta \cdot (1-HBS_{t-k}) \cdot RV_{t-k} + e_t \quad (16)$$

$$Test\ 4: RV_t^2 = \alpha + \beta_1 \cdot HBS_{t-k} + \beta_2 \cdot HBS_{t-k} \cdot RV_{t-k} + \beta_3 \cdot (1 - HBS_{t-k}) \cdot RV_{t-k} + e_t \quad (17)$$

RV_t and HBS_t, respectively, indicate the monthly realized volatility and the housing market instability signal. *Tests 1*, *2*, *3*, and *4* correspond to the sign bias, negative size bias, positive size bias, and joint bias tests, respectively. As with the previous bias tests, we interpret the test results combining the four analyses. Unlike existing bias tests, our test varies by the time difference and checks whether volatility responds asymmetrically to the housing market bubble signal in the long term. Table 2 summarizes the bias test results.

Table 2 shows that stock market volatility moves asymmetrically in the wake of instability signals in the housing market. Furthermore, the coefficients for *Test 1* are significantly negative, and the coefficients for *Test 3* are significantly positive, whereas the coefficients are not significant in *Test 2*. Thus, the housing market bubble significantly impacts stock market volatility. Through these results, we confirm that the EWS proposed in this study can detect increased instability in the stock market as well as in the housing market.

IV. HOUSING MARKET EARLY WARNING SYSTEM

A. MODEL

This section builds an EWS using an LSTM model to predict *HBS* and stock market volatility. The LSTM network

TABLE 3. Confusion matrix for housing market early warning.

Actual \ Predicted	1: Bubble	0: Non-Bubble
1: Bubble	True Positive (TP)	False Negative (FN)
0: Non-Bubble	False Positive (FP)	True Negative (TN)

is a sequential forecasting method by learning long- and short-term dependencies [35], [36]. It enables the processing of sequential data with arbitrary lengths via a hidden state vector and also improves the learning power of long-distance dependency by introducing a so-called memory cell.

We design an LSTM network following Wang *et al.* [37]. The inputs to an LSTM cell at time t are the hidden state (h_{t-1}) and the cell state (C_{t-1}) that contain historical information from the former cell. We employ three gates; the forget, input, and output gates determine the information to be discarded, added, reproduced, respectively. The sigmoid functions for the forget (Γ_f), input (Γ_i), and output (Γ_o) gates each reflect the information level that the given gate processes after adjusting the current input and the previous hidden state. $\tilde{C}_t$ is the new candidate state created by the *tanh* layer, which is added to the cell state, C_{t-1}, to generate the next cell state, C_t. The formulae of the three gates and the new candidate state can be written as:

$$\begin{aligned}
\Gamma_f &= \sigma\left(x_t U^f + h_{t-1} W^f + b_f\right); \\
\Gamma_i &= \sigma\left(x_t U^i + h_{t-1} W^i + b_i\right); \\
\Gamma_o &= \sigma\left(x_t U^o + h_{t-1} W^o + b_o\right); \\
\tilde{C}_t &= tanh\left(x_t U^g + h_{t-1} W^g + b_g\right),
\end{aligned} \tag{18}$$

where σ is the sigmoid function, x_t is the input vector, h_{t-1} is the hidden state, U and W are the weighted matrices connecting the current layer with the input vector and previous layer, respectively, and b is the bias. Finally, an LSTM cell at time t produces three outputs: the output vector that includes HBS and stock market volatility, the next hidden state (h_t), and the next cell state (C_t). h_t and C_t recur as the inputs to the next cell.

We now evaluate the housing bubble predictor based on the LSTM neural network in comparison with those based on our two baseline models: the random forest (RF) and support vector machine (SVM) models. For the evaluation of the predictor, we employ the Rand accuracy metric, which is designed for classification models [38]. The confusion matrix for the evaluation is shown in Table 3.

The true positive (TP) and negative (TN) correspond to true predictions, whereas the false positive (FP) and negative (FN) correspond to false predictions [39], [40]. The Rand accuracy is defined as the percent of true results:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{19}$$

In Equation (19), a higher Rand accuracy indicates a greater level of predictive power.

TABLE 4. Descriptive statistics of the input variables.

	Mean	Std.	Min	Max
ps	0.0823	0.196	-0.4609	0.5579
er	0.0063	0.121	-0.2435	0.633
rr3	-0.0436	0.21	-0.4235	0.5427
m3	0.0796	0.016	-0.1677	0.1314
dep	0.0917	0.213	-0.3788	0.719
lr	0.0376	0.016	0	0.0919
dlr	0.001	0.015	-0.0476	0.05
cpi	0.0217	0.012	-0.0043	0.059
ipi	0.0289	0.034	-0.0861	0.1387
pscon	0.0951	0.398	-0.6552	1.5985
contl	0.112	0.402	-0.6349	1.8814
has	0.1034	0.502	-0.7882	3.3259
tlarea	0.0381	0.217	-0.5769	1.0911

B. DATA

The input variables used in this study are as follows. The monthly macroeconomic variables, collected from November 2004 to December 2019, include the KOSPI index (*ps*), the USD/KRW exchange rate (*er*), the interest rate on three-year treasury bonds (*rr3*), total liquidity (*m3*), customer deposits (*dep*), the composite leading economic index (*lr*), the loan-to-deposit rate (*dlr*), the consumer price index (*cpi*), and the industrial production index (*ipi*). The real estate market variables include the construction stock index (*pscon*), large-scale construction orders (*contl*), apartment supply quantity (*has*), and the total area of transacted land in residential areas (*tlarea*). All variables are collected as the year-on-year rate of changes, except *m3*, *rr3*, and *dlr*, which are provided as the relative changes to the values one month, three months, and two months prior, respectively. Table 4 summarizes the descriptive statistics of the variables.

C. RESULTS

This study uses an LSTM neural network and constructs a model that predicts values of housing and stock market variables at time t based on the values at time $t-1$. We use 70% of the full sample as training data and the remaining 30% as forecasting data. Thus, for $T = 171$, observations at $t = 1, \ldots, 120$ are used as the training data set, and observations at $t = 121, \ldots, 171$ are used as the test data.

We develop three models. The first model (*Model 1*) uses *HMPI* and monthly volatility as input variables. The second model (*Model 2*) uses *HBS* instead of *HMPI*. The last model (*Model 3*) uses both *HMPI* and *HBS*. Predictions are made from a network of LSTM memory blocks that sequentially process the input variables to each model. The batch size and the epoch number are 20 and 250, respectively. We use the adaptive moment estimation (Adam) optimizer and set the initial learning and dropout rates as 0.005 and 0.02, respectively. A comparison of the models is shown in Table 5.

As Table 5 shows, *Model 2*, which includes *HBS*, has an overall lower root mean square error for *RV* prediction than that for the prediction in *Model 1*, which does not include *HBS*, has. Moreover, the results for *Model 3* indicate that including *HBS* in addition to *HMPI* improves the prediction accuracy of *RV*. Our findings are also shown in Figure 2.

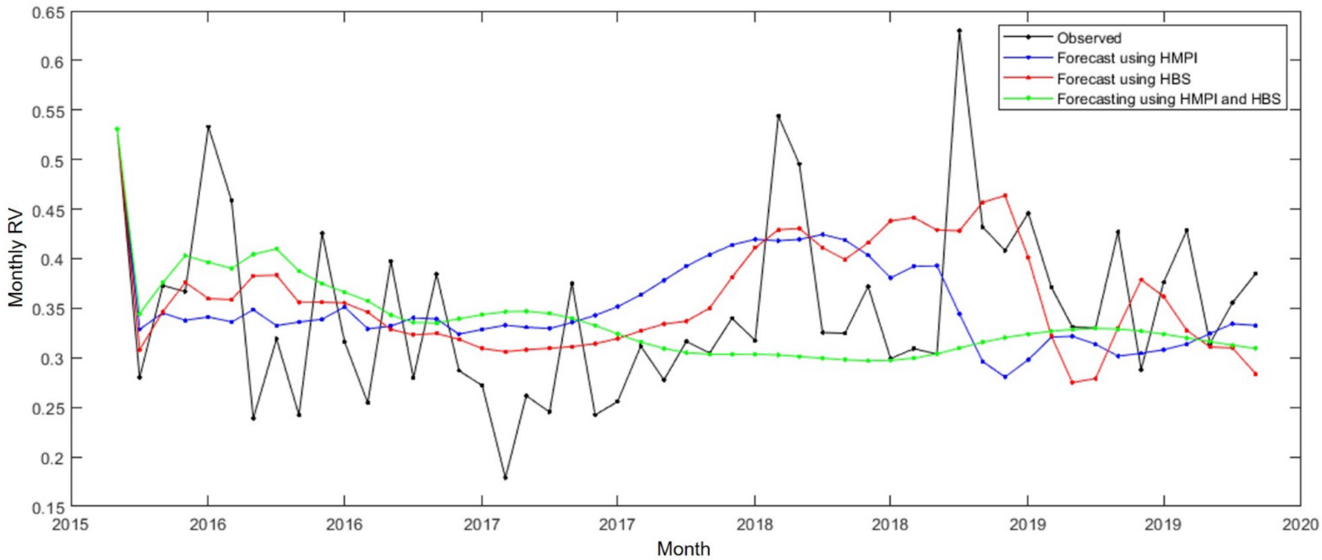

FIGURE 2. Forecasting the stock market volatility.

TABLE 5. Root mean square errors.

	Model 1	Model 2	Model 3
HMPI	127.13		130.05
HBS		1.0239	1.2477
RV	1.0354	1.0213	1.0318
ps	0.9844	0.9749	1.0176
er	0.9783	0.9539	0.9703
rr3	1.0147	1.0233	1.1933
m3	0.9180	0.9155	0.9185
dep	0.8995	0.9207	0.9688
lir	0.9870	0.9868	0.9927
dlr	1.00007	1.0006	1.0000
cpi	0.9872	0.9871	0.9915
Ipi	0.9861	0.9870	0.9876
pscon	1.0128	0.9985	1.0033
contl	0.9634	0.9502	1.0644
has	0.7358	0.7496	0.7176
tlarea	0.9842	0.9780	1.0167

Figure 2 compares the stock market volatility predictions by *Models 1*, *2*, and *3* in comparison to the realized monthly volatility. The figure shows that *Model 2*, which only includes *HBS*, predicts abrupt volatility changes better than other models, and the prediction performance of *Model 2* is better than *Model 3*, which uses *HMPI* and *HBS*. This result means that considering only the housing market bubble signals is more suitable for predicting the stock market volatility than considering housing market price data as well.

We now compare the crisis prediction of the LSTM model with those of the RF and SVM models using the evaluation metrics. In the classification model, the batch size and the epoch number are 27 and 60, respectively. The Adam optimizer is used. The initial learning and dropout rates are 0.001 and 0.1, respectively. Table 6 lists the Rand accuracies of the three models, using the test-set data.

TABLE 6. Accuracies of LSTM, RF, and SVM models.

	LSTM	RF	SVM
Accuracy	0.9915	0.9316	0.9573

The results in Table 6 suggest that the LSTM model produces the optimal crisis prediction, as it yields the greatest accuracy (0.9915) among all of the examined cases.

V. CONCLUSION

Housing market bubbles damage the overall national economy. Especially, the damage from housing market-driven financial crises is getting enormous. The response to the crisis is more likely to be effective if both the housing market bubbles and financial market crises can be identified simultaneously in advance. Accordingly, this study develops an EWS to detect instability in both the housing and stock markets. We identify the housing market bubble signal and predict the signal using the LSTM neural network. By demonstrating that the signal affects stock market volatility asymmetrically, we show that our EWS can comprehensively detect risk in the housing and stock markets. This result highlights the better practicality of our EWS in response to financial crises over the existing EWSs.

This study has an academic contribution in that it constructs an effective EWS by uniquely defining housing bubble signals and reveals the link between these signals and financial market volatility. By proposing a comprehensive EWS, we offer policy implications regarding risk in both the housing and stock markets. Meanwhile, since the predictive power of this EWS is vulnerable to drastic policy changes, further extension of the model to include policy and macroeconomic variables may be necessary for future studies.

REFERENCES

[1] I.-H. Cheng, S. Raina, and W. Xiong, "Wall street and the housing bubble," *Amer. Econ. Rev.*, vol. 104, no. 9, pp. 2797–2829, Sep. 2014.

[2] J. F. McDonald and H. H. Stokes, "Monetary policy and the housing bubble," *J. Real Estate Finance Econ.*, vol. 46, no. 3, pp. 437–451, Apr. 2013.

[3] H. Kim, H. Cho, and D. Ryu, "Characteristics of mortgage terminations: An analysis of a loan-level dataset," *J. Real Estate Finance Econ.*, vol. 57, no. 4, pp. 647–676, Nov. 2018.

[4] S. I. Seok, H. Cho, and D. Ryu, "The information content of funds from operations and net income in real estate investment trusts," *North Amer. J. Econ. Finance*, vol. 51, Jan. 2020, Art. no. 101063.

[5] J. Clayton, "Rational expectations, market fundamentals and housing price volatility," *Real Estate Econ.*, vol. 24, no. 4, pp. 441–470, Dec. 1996.

[6] D. Fernández-Kranz and M. T. Hon, "A cross-section analysis of the income elasticity of housing demand in Spain: Is there a real estate bubble?" *J. Real Estate Finance Econ.*, vol. 32, no. 4, pp. 449–470, Apr. 2006.

[7] O. Nneji, C. Brooks, and C. Ward, "Intrinsic and rational speculative bubbles in the U.S. housing market: 1960–2011," *J. Real Estate Res.*, vol. 35, no. 2, pp. 121–152, Jan. 2013.

[8] D. Cho and S. Ma, "Dynamic relationship between housing values and interest rates in the Korean housing market," *J. Real Estate Finance Econ.*, vol. 32, no. 2, pp. 169–184, Mar. 2006.

[9] M. Hwang, J. M. Quigley, and J.-Y. Son, "The dividend pricing model: New evidence from the Korean housing market," *J. Real Estate Finance Econ.*, vol. 32, no. 3, pp. 205–228, May 2006.

[10] D. Hagemann and M. Wohlmann, "An early warning system to identify house price bubbles," *J. Eur. Real Estate Res.*, vol. 12, no. 3, pp. 291–310, Nov. 2019.

[11] D. Beckmann, L. Menkhoff, and K. Sawischlewski, "Robust lessons about practical early warning systems," *J. Policy Model.*, vol. 28, no. 2, pp. 163–193, Feb. 2006.

[12] P. C. Phillips, S. Shi, and J. Yu, "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500," *Int. Econ. Rev.*, vol. 56, no. 4, pp. 1043–1078, Oct. 2015.

[13] G. L. Kaminsky, J. S. Lizondo, and C. M. Reinhart, "Leading indicators of currency crises," *IMF Staff Papers*, vol. 45, no. 1, pp. 1–48, 1998.

[14] J. A. Frankel and A. K. Rose, "Currency crashes in emerging markets: An empirical treatment," *J. Int. Econ.*, vol. 41, nos. 3–4, pp. 351–366, Nov. 1996.

[15] L. Alessi and C. Detken, "Identifying excessive credit growth and leverage," *J. Financial Stability*, vol. 35, pp. 215–225, Apr. 2018.

[16] I. Aldasoro, C. E. V. Borio, and M. Drehmann, "Early warning indicators of banking crises: Expanding the family," *BIS Quart. Rev.*, pp. 29–45, Mar. 2018. [Online]. Available: https://www.bis.org/publ/qtrpdf/r_qt1803e.htm

[17] C. Dreger and K. Kholodilin, "An early warning system to predict speculative house price bubbles," *Econ., Open-Access, Open-Assessment E-J.*, vol. 7, pp. 1–26, Mar. 2013.

[18] S. Ferrari, M. Pirovano, and W. Cornacchia, "Identifying early warning indicators for real estate-related banking crises," Eur. Systemic Risk Board, Frankfurt, Germany, ESRB Occasional Paper, Tech. Rep. 2015/08, Aug. 2015, pp. 1–69, no. 8.

[19] J. Cao, Z. Li, and J. Li, "Financial time series forecasting model based on CEEMDAN and LSTM," *Phys. A, Stat. Mech. Appl.*, vol. 519, pp. 127–139, Apr. 2019.

[20] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018.

[21] H. Kim, H. Cho, and D. Ryu, "Corporate default predictions using machine learning: Literature review," *Sustainability*, vol. 12, no. 16, p. 6325, Aug. 2020.

[22] Y. Baek and H. Y. Kim, "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module," *Expert Syst. Appl.*, vol. 113, pp. 457–480, Dec. 2018.

[23] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 2823–2824.

[24] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1419–1426.

[25] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, pp. 25–37, Aug. 2018.

[26] Y. Zhou, R. Han, Q. Xu, Q. Jiang, and W. Zhang, "Long short-term memory networks for CSI300 volatility prediction with Baidu search volume," *Concurrency Comput., Pract. Exper.*, vol. 31, no. 10, p. e4721, May 2019.

[27] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Appl. Soft Comput.*, vol. 90, May 2020, Art. no. 106181.

[28] N.-K. Chen, "Asset price fluctuations in Taiwan: Evidence from stock and real estate prices 1973 to 1992," *J. Asian Econ.*, vol. 12, no. 2, pp. 215–232, Jun. 2001.

[29] J. Okunev, P. Wilson, and R. Zurbruegg, "The causal relationship between real estate and stock markets," *J. Real Estate Finance Econ.*, vol. 21, no. 3, pp. 251–261, Nov. 2000.

[30] H. M. Kim, "Ethnic connections, foreign housing investment and locality: A case study of Seoul," *Int. J. Housing Policy*, vol. 17, no. 1, pp. 120–144, Jan. 2017.

[31] D. Hyun and S. Milcheva, "Spatial dependence in apartment transaction prices during boom and bust," *Regional Sci. Urban Econ.*, vol. 68, pp. 36–45, Jan. 2018.

[32] K. H. Al-Yahyaee, W. Mensi, H.-U. Ko, M. Caporin, and S. H. Kang, "Is the Korean housing market following gangnam style?" *Empirical Econ.*, pp. 1–32, Aug. 2020, doi: 10.1007/s00181-020-01931-2.

[33] S. H. Suh, K. Kim, and J. S. Jeon, "Housing market early warning system: The case of Korea," *Eur. J. Sci. Res.*, vol. 56, no. 4, pp. 539–547, Jul. 2011.

[34] R. F. Engle and V. K. Ng, "Measuring and testing the impact of news on volatility," *J. Finance*, vol. 48, no. 5, pp. 1749–1778, Dec. 1993.

[35] S. Hochreiter and J. Schmidhuber, "LSTM can solve hard long time lag problems," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 473–479.

[36] M. I. Jordan, "Serial order: A parallel distributed processing approach," *Adv. Psychol.*, vol. 121, pp. 471–495, May 1997.

[37] P. Wang, L. Zong, and Y. Ma, "An integrated early warning system for stock market turbulence," *Expert Syst. Appl.*, vol. 153, Sep. 2020, Art. no. 113463.

[38] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.

[39] C. E. Metz, "Basic principles of ROC analysis," *Seminars Nucl. Med.*, vol. 8, no. 4, pp. 283–298, Oct. 1978.

[40] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

DAEHYEON PARK is currently pursuing the Ph.D. degree with the College of Economics, Sungkyunkwan University. By carrying out in-depth policy research on the Korean economy and market, he has many publications in top-tier KCI journals. His current research interests include fintech, big data analysis, and machine learning.

DOOJIN RYU graduated from the School of Electrical Engineering, Seoul National University. He received the Ph.D. degree in management engineering from KAIST. He is currently a Tenured/Full Professor with the College of Economics, Sungkyunkwan University. He was the winner of the first Elsevier Young Researcher Award. He has ranked fourth in the world (Business/Finance) for the two consecutive years, from 2018 to 2019, Journal Citation Reports-Clarivate Analytics. He is currently an Editor of *Investment Analysts Journal* (SSCI) and the Subject Editor of *Emerging Markets Review* (SSCI), *Journal of Multinational Financial Management* (SSCI), and *Emerging Markets Finance and Trade* (SSCI).