Research Article

The Behavioral Mechanism and Forecasting of Beijing Housing Prices from a Multiscale Perspective

Yan Li,[1] Zhaoyang Xiang ⬤,[2] and Tao Xiong ⬤[2]

[1]College of Cultural Management, Wuhan University of Communication, Wuhan 430205, China
[2]College of Economics and Management, Huazhong Agricultural University, Wuhan 430070, China

Correspondence should be addressed to Zhaoyang Xiang; zhyx@mail.hzau.edu.cn

The first-hand house price in Beijing, the capital of China, has skyrocketed with 43 percent annual growth from 2005 to 2017, exerting tremendous adverse effects on people's livelihood and the development of real estate. Thus, exploring the behavioral mechanism and accurate forecasts of house prices is a critical element in making decisions under uncertain conditions and is of great practical significance for both participants and policymakers in real estate. According to the complex features of house price, including nonlinear, nonstationary, and multiscale, and considering the remarkable time and frequency discrimination capability of multiscale analysis in dealing with house price problems, we develop an ensemble empirical mode decomposition- (EEMD-) based multiscale analysis paradigm to investigate the behavioral mechanism and then obtain accurate forecasts of house prices. Specifically, the monthly house price in Beijing over the period January 2005 to November 2018 is first decomposed into several different time-scale intrinsic-mode functions (IMFs) and a residual via EEMD, revealing some interesting characteristics in house price volatility. Then, we compose the IMFs and residual into three components caused by normal market disequilibrium, extreme events, and the economic environment using the fine-to-coarse reconstruction algorithm. Finally, we propose an improved hybrid prediction model for forecasting house prices. Our experimental results show that the proposed multiscale analysis paradigm is able to clearly reveal the behavioral mechanism hidden in the original house price. More importantly, the mean absolute percentage errors (MAPEs) of the proposed EEMD-based hybrid approach are 5.62%, 7.24%, and 8.63% for one-, three-, and six-step-ahead prediction, respectively, consistently lower than the MAPE of the three competitors.

1. Introduction

Reforms in China's system of urban housing, an important part of the "reform and opening up" policy initiated in 1978, led to a general and significant improvement in accommodations for most of the urban population in the country. In nearly four decades, the central and local governments have successfully provided new and owner-occupied and indemnification housing of reasonable quality to as much as 80% of the urban population in China. At the same time, real estate has been a vital engine of rapid growth in China over the past two decades. According to data from the National Bureau Statistics (NBS) of China, real estate investment grew from about 4% of the gross domestic product (GDP) in 1997 to 13.27% of GDP in 2017. In 2017, residential investment in

ground-up developments, in particular, reached a record high of RMB 0.751 trillion (US$111.52 billion), up to 9.4% from RMB 0.687 trillion in 2016. It accounts for both about 68.4% of real estate investment and 11.89% of fixed-asset investment in 2017, which was high compared with that in other developed and developing countries. Real estate has strong linkages to several upstream and downstream industries, and sales are also a key source of local public finance. Therefore, healthy growth in real estate is important in economic development in China.

Beijing house price has exhibited considerable soar over the past two decades. According to NBS data, the first-hand (new) house price in Beijing ranged from RMB 8,332 (US$1,200) per sq.m. (square meter) in January 2005 to RMB 52,405 (US$7,600) per sq.m. in July 2017. Thus, criticisms of

residential housing investment centered on overpricing, particularly in central Beijing, where many properties were sold at prices exceeding RMB 100,000 (US$17,000) per sq.m. On March 15, 2018, a survey commissioned by the People's Bank of China (PBOC, the central bank) showed that 62.9% of urban residents in a sample of 50 cities thought that house prices had become "unacceptably high." This sentiment has risen by eight percentage points or more every quarter since 2010, especially among high- and middle-income earners. In addition, 31.4% of urban residents expected the future trend in house prices will continue to accelerate, while 48.2% of them expected it to remain stable; only 9.9% of urban residents expected it to decline (this survey was commissioned by the People's Bank of China (PBOC) in 2018 in which the respondents are sampled randomly in fifty cities; in this survey, the interviewer could not distinguish permanent urban residents from those who were transitory, so the concept of "urban residents" used in this survey includes all urban residents).

The extraordinary rise in housing prices in China has increased the risk and uncertainty faced by central and local governments when making decisions related to land auctions, mortgage policy, land supply, monetary policy, and fiscal policy. A high house price is detrimental to consumer welfare, but low house prices are harmful to government revenue and prompt it to make some decisions, such as reducing the supply of salable land, which further worsens volatility in house prices in the future (Wen and Goodman [1] concluded that the rising housing price causes land price to increase and vice versa; thus, declining housing price causes land price to decrease, which ultimately results in the lower government revenue, taking the existence of so-called land public finance at the local level into account). Therefore, an understanding of behavioral mechanism and accurate forecasts of house prices is a critical element of making decisions under uncertain conditions that significantly affect consumer welfare and government revenue and is of great practical significance for policymakers when designing real estate regulations.

The existing literature demonstrates many efforts to analyze real estate pricing behavior ([1–12]; Wu, 2015). For example, Hott and Monnin [7] proposed two alternative models, based on a no-arbitrage condition between renting and buying and the market equilibrium between housing demand and supply, respectively, to estimate fundamental prices on real estate markets. Hossain and Latif [8] examined the determinants of housing price volatility and investigated the dynamic effects of these determinants on volatility, by using GARCH and VAR models. Ren et al. (2012) applied the theory of rational expectation bubbles to the Chinese housing market. Based on data in 35 cities in China, they found no evidence of such bubbles in the Chinese housing market. Du et al. [4] examined the impacts of land policies on the dynamic relationship between housing and land prices in the Chinese real estate market and found that a unidirectional Granger causality between housing and land prices exists in the short run. Gupta et al. [6] assessed the impact of monetary policy on inflation in house prices in the nine census divisions of the US economy using a factor-

augmented vector autoregression (FAVAR) model. Wen and Zhang (2013) developed a simultaneous equations model to explore the interaction between housing prices and land and found that the rising housing price causes land prices to increase and vice versa. Zhang et al. [12] investigated the effect of regulations on the relationship between housing prices and volume in China. Wang and Xie [10] examined the correlation structure and dynamics of international real estate securities markets using a minimum spanning tree, the hierarchical tree, and the planar maximally filtered graph.

Many traditional statistical techniques have been used in real estate price forecasting, such as autoregressive model [13], dynamic model averaging [14], lattice models and geostatistical models [15], and fixed-effects model [16]. For example, Clapp and Giaccotto [13] used an autoregressive process, which was adopted to produce one-quarter-ahead forecasts for individual properties, to model a city-wide housing price index. Beracha et al. [16] examined the extent to which future cross-sectional differences in house price changes were predicted based on online search intensity in prior periods. Bork and Moller [14] examined house price predictability across the US using dynamic model averaging and dynamic model selection, which allows for model change and parameter shifts. Under the assumption of linear hypothesis, these econometric models perform well. However, due to the significant nonlinear hidden patterns in the real estate price, those statistical and econometric models are not suitable for forecasting real estate prices under real-life conditions.

Therefore, the research on machine learning has recently emerged as a trend in the field of real estate price forecasting because it can capture the nonlinearity and nonstationarity that exist widely in real estate prices. Specifically, support vector regression [17], C4.5 [18], Naïve Bayesian [18], AdaBoost [18], Boltzmann machine [19], and grey models [20] have been wildly used to forecast house prices. For instance, Wang et al. [17] proposed a hybrid model that integrates support vector machines and particle swarm optimization for real estate price forecasting. Rafiei and Adeli [19] developed a machine learning-based comprehensive model for estimating the price of new housing in any given city at the design phase or the beginning of the construction. Although the above studies have shown that machine learning techniques can effectively analyze complex systems, machine learning techniques also have their own limitations, such as parameter sensitivity, local optimization, and overfitting [21].

Hybrid models have attracted increasing interest in time series analysis and prediction. A hybrid model makes full use of the strengths of different methods to make up for the shortcomings of other methods to improve the analytical performance of the model [21]. Among the hybrid models, the multiscale analysis based on the empirical mode decomposition (EMD) or ensemble empirical mode decomposition (EEMD) is a cutting-edge technology in the study of time series analysis and prediction. The main aim of the multiscale analysis is to interpret the generation of time series from a novel perspective [22]. Previous studies in

which real estate prices are treated as a single series (original scale) have difficulty in analyzing internal driving forces and their economic meaning for movements in house prices and thus have poor performance in forecasting house prices. By decomposing the real estate price into intrinsic-mode functions (IMFs), the original tough analysis and forecasting task could be divided into several relatively easy subtasks, particularly taking the high complexity and irregularity of real estate prices into account. As we propose, the multiscale analysis paradigm is a promising alternative that can partially solve this problem.

Our focus in this study is on the behavioral mechanism and forecasting of house prices in Beijing from a multiscale perspective. To our knowledge, this study is the first to use a multiscale analysis of real estate with an application of the EEMD technique. In this way, we add to a fairly limited body of research in this field. On the basis of empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD), multiscale analysis is a promising approach for deeply exploring the behavioral mechanism and forecasting of a price series with several scales, rather than the original scale. Successful examples of applications are seen in various fields, including but not limited to the energy market [22–24], the financial market [25, 26], the carbon market [27, 28], and signal processing [29]. EMD, first proposed by Huang et al. [30], changes a local and high-adaptive decomposition of a time-series into intrinsic-mode functions (IMFs) with different average time scales. The main advantage of EMD is its ability to handle nonlinear processes because this technique has no a priori assumptions regarding these properties of price series under consideration. Still, each of the IMFs obtained reflects the dynamics of the price series at a specific time scale, which allows studying the fine structure of the price series [23].

Inspired by multiscale analysis, this study develops a multiscale analysis framework based on EEMD [31] and a fine-to-coarse reconstruction algorithm [22] to explore the behavioral mechanism of house prices in Beijing. In particular, we reexamine the periodicity and nonlinearity of and the effects of regulations on house prices from a multiscale perspective. Furthermore, we propose an EEMD-based hybrid approach to perform short-term forecasting of house prices in Beijing. The monthly price of a house in Beijing from January 2005 to November 2018 is used as experimental data for the purpose of validation. In this multiscale analysis paradigm, the original house price series is first decomposed using EEMD, which is a substantial improvement over the original EMD, into several IMFs and a residual. Then, we employ the fine-to-coarse reconstruction algorithm to compose the IMFs obtained and residual into high-frequency, low-frequency, and trend components, which have their own economic meaning. Specifically, the economic meaning of these three components is identified as short-term fluctuations originating from normal market disequilibrium of supply and demand, the effects of extreme events, such as a financial crisis and the release of regulations, and a long-term trend, respectively. By doing so, we can examine the characteristics and underlying rules of house prices. From the perspective of multiscale analysis, we further propose a four-step modeling framework, integrating EEMD, fine-to-coarse reconstruction algorithm, autoregressive integrated moving average (ARIMA), polynomial function, and support vector regression (SVR), for forecasting short-term house prices.

The rest of this paper is organized as follows. Section 2 explains the EEMD, fine-to-coarse reconstruction algorithm, and the proposed EEMD-based hybrid perdition model. Section 3 introduces the datasets used. Section 4 details and discusses our decomposition results. Finally, Section 5 concludes the paper.

2. Theoretical Background and Methodologies

2.1. Ensemble Empirical Mode Decomposition (EEMD). Ensemble empirical mode decomposition (EEMD) was proposed by Wu and Huang [31] recently to solve the mode mixing problem existing in the original empirical mode decomposition (EMD) by adding white noise. First of all, a mixed mode is defined as consisting of a single "intrinsic-mode function (IMF)" that comprises either signals of widely disparate scales or a single similar scale in different IMF components [31]. Adding white noise to the original house price series can significantly limit the problem of a mixed mode in the original EMD method. EEMD is based on the insight gleaned from recent studies of the statistical properties of white noise (Flandrin et al., 2004; [31]), which showed that the EMD is effectively an adaptive dyadic filter bank when applied to white noise. As explained by Wu and Huang [31], the principle of the EEMD is as follows: the added white noise populates the entire time-frequency space uniformly with constituent components of different scales. Because the noise in each trial is different, it is canceled out in the ensemble mean of enough trials. The ensemble mean is treated as the true answer.

Huang et al. [30] used the term "intrinsic-mode function" because it represents the oscillation mode embedded in the data. The IMF in each cycle, defined by the zero crossings, involves only one mode of oscillation, with no complex riding waves allowed. Note that the time scale used in this study is the inverse of the frequency for time series in general. In EEMD, the smaller the time scale, the more "compressed" or higher frequency the IMFs. Conversely, the larger the time scale, the more "stretched" or lower frequency the IMFs.

EEMD can decompose a complex time series into a set of IMFs as well as a residual, which reveals the oscillation mode embedded in the time series. Since the work of Wu and Huang [31], EEMD has been successfully used to analyze time series with complex nonlinearity and high irregularity in various areas. Theoretically, IMFs must meet the following two conditions: (1) the number of extremes and of zero crossings must be equal or must differ at most by one; (2) the time series must be symmetric with respect to local zero means.

Given a house price series X_t for $t = 1, 2, \ldots, n$, according to this definition of IMFs, IMFs can be extracted iteratively from this series using the following sifting process:

Step 1: identify all local extremes (both maxima and minima) in the house price series.

Step 2: connect all local maxima and minima using cubic spline interpolation to generate its upper and lower envelope lines, X_t^{upper} and X_t^{lower}, respectively.

Step 3: compute the pointwise local mean $m_t = (X_t^{\text{upper}} + X_t^{\text{lower}})/2$ from the upper and lower envelope lines.

Step 4: define the difference between the house price series X_t and the mean m_t: $h_t = X_t - m_t$.

Step 5: examine whether h_t satisfies the two aforementioned conditions of IMFs. If so, h_t is a new IMF, so replace X_t with $r_t = X_t - h_t$; if not, replace X_t with h_t.

Step 6: repeat Steps 1 to 5 until the stop criterion is satisfied.

After this sifting process is completed, the original house price series $X(t)$ can finally be expressed as the sum of IMFs and a residual is extracted:

$$X_t = \sum_{i=1}^{k} h_{i,t} + r_t, \quad t = 1, 2, \ldots, n, \tag{1}$$

where k denotes the total number of IMFs, $h_{i,t}$ $(i = 1, 2, \ldots, k)$ are the IMFs, and r_t is the residual.

Adding white noise to the original house price series can significantly reduce the mixed-mode problem in the original EMD method. As such, the EEMD procedure is presented as follows:

Step 1: add a white noise series ε_t to the original house price series X_t to generate a new price series $X_t^{\text{new}} = \varepsilon_t + X_t$ for $t = 1, 2, \ldots, n$.

Step 2: decompose the new price series X_t^{new} into a set of IMFs as well as a residual via the sifting process outlined above.

Step 3: repeat Steps 1 and 2 using different white noise series each time. Finally, calculate the ensemble means of corresponding IMFs and residuals, which generates the final results.

2.2. Fine-to-Coarse Reconstruction Algorithm. As discussed above, the original house prices are decomposed into a set of IMFs and a residual via EEMD. Basically, EEMD can be applied as a filter to separate the high-frequency patterns from the low-frequency ones. To facilitate the prediction modeling and reveal the economic meaning of the decomposition results via EEMD, we use the fine-to-coarse reconstruction algorithm [22], a high-pass filter, to reconstruct the IMFs by adding fast oscillations to slow ones and then generate three components: high-frequency, low-frequency, and trend components. Given a set of IMFs $h_{i,t}$ $(i = 1, 2, \ldots, k)$ of house prices obtained in the previous section, we describe the fine-coarse reconstruction algorithm as follows [22]:

Step 1: calculate the mean of the sum of $h_{1,t}$ to $h_{r,t}$, $r = 1, 2, \ldots, k$, which produces M_r $(r = 1, 2, \ldots, k)$.

Step 2: use a t-test to determine for which $r (r = 1, 2, \ldots, k)$ the mean M_r significantly departs from zero at the significance level of 0.05.

Step 3: after r is determined, the IMFs from r to k are partially summed as the low-frequency components, and the rest of the IMFs are partially summed as the high-frequency components. Moreover, the residual is treated as the trend component.

2.3. The Proposed EEMD-Based Hybrid Prediction Approach. To obtain an accurate prediction of house prices in Beijing, we proposed an EEMD-based hybrid approach and present the steps involved in this process in this section. Given a housing price series $x(t)$ for $t = 1, \ldots, n$, we suggest six-month-ahead forecasting. We construct a four-step model framework for short-term house price forecasting that integrates EEMD, fine-to-coarse reconstruction algorithm, ARIMA, polynomial function, and support vector regression (SVR) (see Figure 1).

As seen in Figure 1, the proposed EEMD-based hybrid approach consists of the following four steps:

Step 1: decomposition. The original house price is decomposed into several IMFs and a residual using the EEMD technique.

Step 2: composition. The fine-to-coarse reconstruction algorithm is applied to compose the obtained IMFs and residual into high-frequency, low-frequency, and trend components.

Step 3: single forecasting. SVR is used for low-frequency component forecasting, while ARIMA and polynomial function are used to forecast the high-frequency and trend components, respectively.

Step 4: ensemble forecasting. The predicted values of the high-frequency, low-frequency, and trend components are aggregated using another independent SVR model, which models the relationship among the three parts, to generate an ensemble forecast—the final prediction of the original house price series.

3. Datasets

In this study, we use monthly data on the price of new residential houses in Beijing obtained from the NBS. Beijing is the capital of China, the second most populous city and most populous capital city in the world. The monthly price is the average transaction price in nominal terms across all districts in Beijing, which is recorded by the city's Department of Housing Management. Figure 2 shows the house price series in Beijing from January 2005 to November 2018, with a total of 167 data points as well as some significant events related to the real estate market and when they occurred.

Figure 2 illustrates that house prices have had extraordinary fluctuations and a steady uptrend since 2005. Based on the time gap between stationary phases in housing prices, an obvious periodicity of approximately three years emerges. The lowest price in our sample (RMB 8,332 per sq.m.)
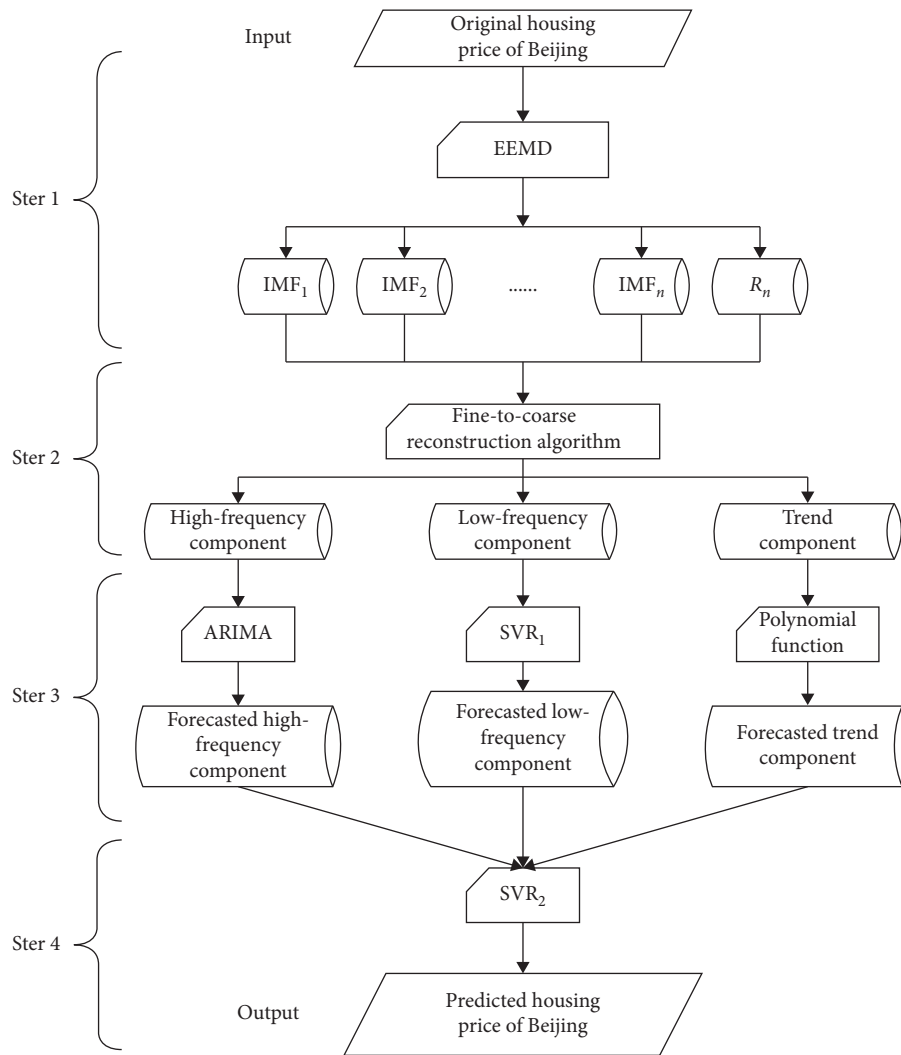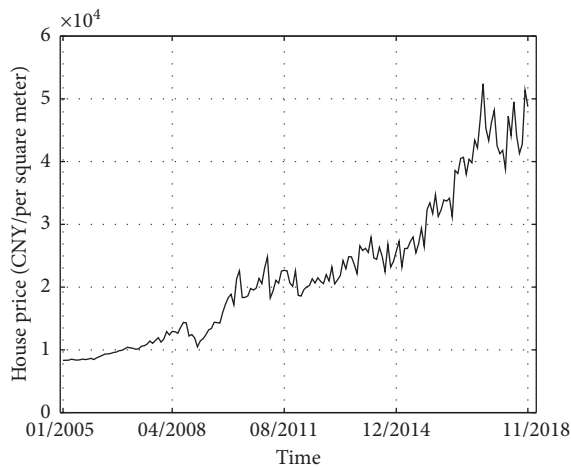
Figure 1: The proposed EEMD-based hybrid approach.

—— House price of Beijing (CNY/per square meter)

Figure 2: Monthly house prices in Beijing.

occurred in January 2005, and then the price rose to RMB 14,308 per sq.m. in August 2008. Afterward, the price remained stable until November 2009, before sharply rising from RMB 18,306 per sq.m. in December 2009 to RMB 21,078 per sq.m. in May 2011, followed by another stationary phase from May 2011 to December 2012. In March 2014, the price reached its second-highest value (RMB 27,913 per sq.m.). Since mid-2014, alternating phases of stability and increases have followed, such that in June 2017, the price attained its highest value to date (RMB 52,405 per sq.m.). Because of a series of strict central and local government regulations since then, the price stopped rising and entered another phase of stationary fluctuation.

Table 1 lists the summary statistics of house prices. The estimated measure of skewness suggests that house prices have a nearly symmetric distribution. In addition, the kurtosis suggests that house prices have lighter tails than a standard normal distribution. The results of the nonnormality Jarque–Bera test at the 1% level of significance further suggest

TABLE 1: Statistical description of the housing price series.

Statistic	Value
Sample size	167
Mean	22,537.220
Std. dev.	11,571.440
Min.	8,332.510
Max.	52,405.000
Median	20983.000
Skewness	0.753
Kurtosis	2.682
Jarque–Bera test	16.489***

Note. ***The null hypothesis is rejected at the 0.01 level (two-tailed).

TABLE 2: Augmented Dickey–Fuller (ADF) test results on the house price series.

Bound	Test for unit root in	T-statistics	p value
ADF	Level	0.657	0.991
	First difference	−13.330***	0.000
Phillips–Perron	Level	0.640	0.990
	First difference	−23.695***	0.000

Note. ***The null hypothesis is rejected at the 0.01 level (two-tailed).

that the null hypothesis of following a normal distribution is rejected for house prices. The augmented Dickey–Fuller (ADF) and Phillips–Perron statistics in Table 2 indicate that a unit root cannot be rejected at the levels. However, house prices are stationary after the first difference.

4. Results and Discussion

4.1. Decomposition. In this section, house prices are decomposed via EEMD into a set of IMFs and a residual. In EEMD, an ensemble membership of 100 is used, and the added white noise in each ensemble member has a standard deviation of 0.2 [32]. Theoretically, the number of IMFs is restricted to $\log_2 N$, where N is the sample size (167 in this case). Therefore, six different time-scale IMFs and a residual are obtained. Figure 3 presents all IMFs and residual listed in order from the highest frequency to the lowest frequency.

All the IMFs present changing frequencies and amplitudes, which is not the same with any harmonic. With the frequency changing from high to low, the amplitudes of the IMFs become larger: for example, all the amplitudes of IMF2 to IMF5 in Figure 3 are up to 3,000; the last residual is the mode slowly increasing from 7,819 to 45,845.

The summary characteristics of the IMFs and residual are reported in Table 3, which shows that the residual accounts for the vast majority of the variance. Indeed, the residual accounts for 94.922% of the total variation in the original house prices, which suggests that they are determined mainly by long-term trends (i.e., the residual). At the same time, the Pearson correlation coefficient and the Kendal coefficient between the original house prices and residual are the highest. In addition, each of the IMFs has a very low correlation coefficient with the original price and accounts for less than 6% of the total variance, further indicating that IMFs have a limited impact on the original prices. An exception occurs in the relationship between the IMF5 and the original prices. When we take a closer look at the IMFs, we find that the second important mode is the lowest frequency IMF, that is, IMF5, which has a mean period of nearly 52.52 months (nearly four years).

4.2. Composition Results. After the analysis of the IMFs and residual is performed, the fine-to-coarse reconstruction algorithm presented above is employed to compose IMFs into high-frequency and low-frequency components. At the same

time, the residual is treated as the trend component. The overview and economic meaning of these three composed components are detailed in this section.

4.2.1. Overview of the Three Composed Components. Table 4 reports the mean of the fine-to-coarse reconstruction as a function of IMF index r, and the house price departs significantly from zero at $r = 4$. Thus, the partial reconstruction with IMF1, IMF2, and IMF3 represents the high-frequency component, and the partial reconstruction with IMF4, IMF5, and IMF6 represents the low-frequency component. The residual is treated as the overall trend in the observed house price.

Figure 4 shows the three composed components of house prices, showing that each component has some distinct features. The high-frequency component reflects short-term fluctuations house prices and thus should be representative of the effects of normal market disequilibrium, which mainly derive from normal volatility in the supply of and demand for commercial houses. By integrating every sharp up or down in the low-frequency component and some extreme events marked in Figure 4, we conclude that each significant break in the low-frequency component corresponds to one or more extreme events and shows their effects. Last but not least, the residual, which slowly varies around the long-term mean, can be treated as the long-term trend during the evolution in house prices.

The statistical measures of these components and the observed price are given in Table 5, which shows that the most dominant component is also the trend component. Indeed, the trend has a high correlation with original house prices, regardless of which correlation statistics are considered. In addition, the low-frequency and trend components have higher contributions to variance, accounting for 3.862% and 94.922%, respectively. Unsurprisingly, the effect of the high-frequency component on all changes in the original housing prices is the smallest.

4.2.2. Long-Term Trend. As mentioned in the previous section, the residual is the component that determines house prices. An important determinant of housing prices is the amount of money in circulation. Figure 5 shows the Chinese money supply (M2) and the trend component obtained in this section, which shows that they have a similar upward trend. In China, therefore, the continuously increasing trend in house prices is consistent with an increase in the money supply.
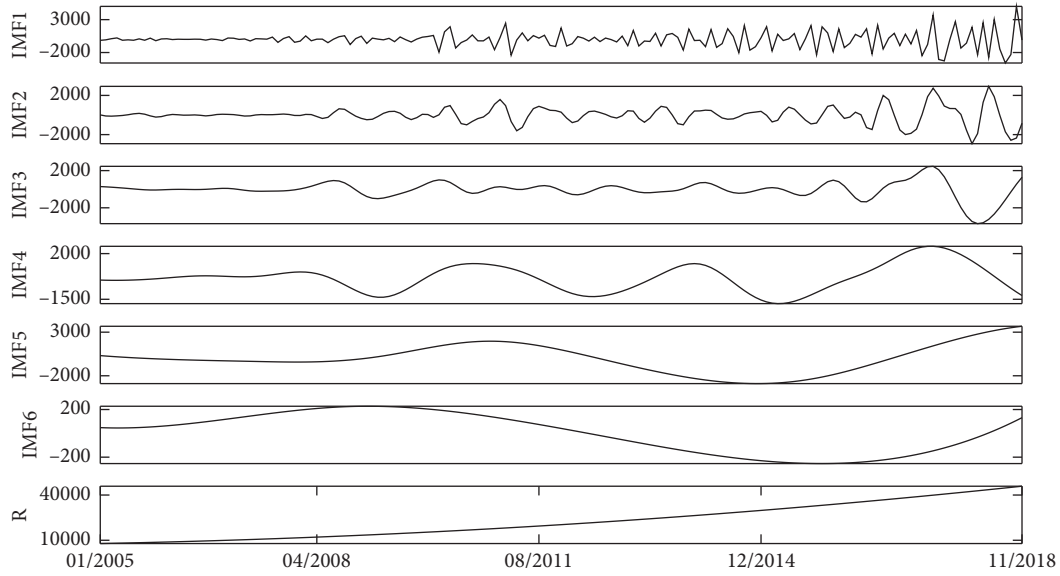
Figure 3: The IMFs and a residual of house prices via EEMD.

Table 3: Measures of IMFs and residual in the house price series.

	Mean period (month)	Pearson correlation	Kendall correlation	Variance	Variance as (%) of observed	Variance as % of ($\sum$ IMFs + residual)
Observed				$1.339 * 10^8$		
IMF1	2.79	0.114	0.061	$1.443 * 10^6$	1.078	1.084%
IMF2	5.95	0.051	0.045	$4.825 * 10^5$	0.360	0.363%
IMF3	15.22	−0.010	0.028	$5.439 * 10^5$	0.406	0.409%
IMF4	33.41	0.281***	0.139**	$9.925 * 10^5$	0.741	0.746%
IMF5	52.52	0.373***	0.021	$2.239 * 10^6$	1.672	1.682%
IMF6	83.23	−0.554***	−0.352***	$2.952 * 10^5$	0.220	0.222%
Residual		0.973***	0.891***	$1.271 * 10^8$	94.922	95.495%
Sum					99.400	100%

Note. *** Correlation is significant at the 0.01 level (two-tailed). **Correlation is significant at the 0.05 level (two-tailed). *Correlation is significant at the 0.1 level (two-tailed).

Table 4: The mean of the fine-to-coarse reconstruction as a function of index r

r	Mean	p-value
1	5.215	0.955
2	−4.924	0.965
3	−31.785	0.137
4	63.252	0.035**

In Table 5, Pearson and Kendall correlation coefficients between the trend component and original house prices are 0.973 and 0.891, respectively, with significance at the 0.01 level. More importantly, the trend component contributes approximately 94.92% of the total variance in original house prices, which suggests that the trend component is the dominant force in the evolution of house prices in the long run. A comparison of the trend component and original housing prices shows that although house prices fluctuated dramatically because of some extreme events, they
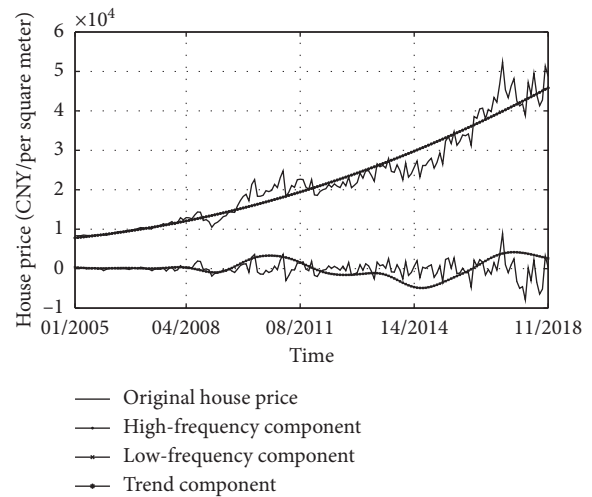

Figure 4: Original house prices and three corresponding composed components.

TABLE 5: Measures of three composed components of house prices.

	Mean period (month)	Pearson correlation	Kendall correlation	Variance	Variance as (%) of observations	Variance as % of ($\sum$ IMFs + residual)
Observed				$1.339 * 10^8$		
High	3.08	0.101	0.062	$3.072 * 10^6$	2.294	2.270%
Low	41.77	0.231***	−0.025	$5.171 * 10^6$	3.862	3.821%
Trend		0.973***	0.891***	$1.271 * 10^8$	94.922	93.910%
Total					101.078	100%

Note. High, low, and trend mean high-frequency, low-frequency, and trend components, respectively. ***Correlation is significant at the 0.01 level (two-tailed).
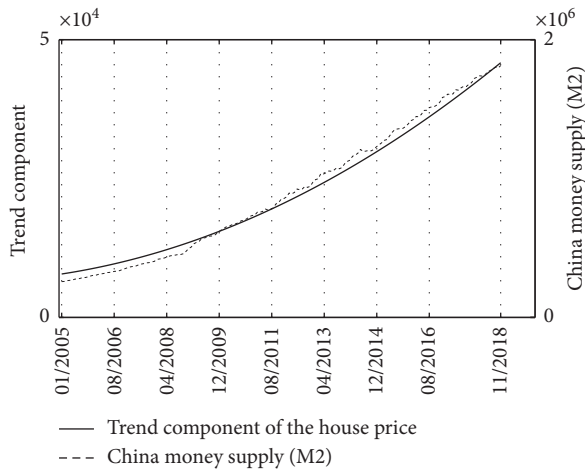

FIGURE 5: Monthly Chinese money supply (M2) and trend component of house prices.

approached and finally returned to the long-term trend after the influence of those extreme events gradually disappeared. For instance, the outbreak of the global financial crisis in mid-2008 made housing prices fall rapidly, from RMB 14,370 to 10,486 per sq.m., but prices recovered afterward and by December 2009 had returned to the trend level, about RMB 18,306 per sq.m.

4.2.3. Effects of Extreme Events. As discussed above, the effects of extreme events on house prices are reflected in the low-frequency component, which consists of IMF4, IMF5, and IMF6. In Table 3, the mean periods of these three IMFs are 33.41 months, 52.52 months, and 83.23 months, respectively, revealing that, historically, the effects of extreme events are often significant and can last for at least three years. In Table 5, the Pearson and Kendall correlation coefficients between the low-frequency component and original house prices are 0.231 and −0.025, respectively. At the same time, the low-frequency component contributes approximately 3.862% of the total variance in original house prices, indicating that the low-frequency component has little influence on changes in house prices, less than that of the trend component. In Figure 4, the low-frequency component is relatively volatile, suggesting that the number and intensity of extreme events in real estate have increased since 2010.

By separating the effects of extreme events contained in the low-frequency component from the original house

prices, we can examine the effect of each extreme event and, more importantly, use it as a benchmark for judging the effect of the next similar extreme event on real estate in Beijing. Figure 6 shows the low-frequency component of the original house prices. More importantly, some significant events related to real estate and when they occurred are indicated in the figure. Then, the policy type, date of release, title, and main focus of the regulations in real estate are detailed in Table 6.

Regulations issued by central and local governments are among the most important events affecting house prices in China [11]. In addition, the changes in local land leasing also significantly affect the housing prices in cities like Beijing [11]. Since 2004, land leases have become an important source of revenue for local governments. Leases generated an amount of revenue equivalent to about 7.5 percent of GDP for Beijing between 2013 and 2017. Basically, local land leasing could be deemed to an event like regulation policy because the auctions on land leasing are not performed every day, and the area and location of land parcels are totally determined by the local government. To illustrate the effect of extreme events on housing prices for the perspective of multiscale analysis, we take the regulation policies as examples in this section.

To better understand the history of policies and the impact of policies on real estate, we briefly review the regulations listed in Table 6. On May 26, 2005, seven ministries released the first official policy to curb the real estate speculation. The policy states that homeowners who sell their house within two years of purchase are subject to business tax on the total revenue from the sale. However, this policy had greater symbolism than practical impact. To further dampen real estate speculation, nine ministries released another policy, which is seen as an improved and enhanced version of the previous policy.

Despite these extensive and severe restrictions, house prices continuously rose. On September 27, 2007, the central government responded with additional measures. For instance, the down payment was raised to 40% and interest rates were increased to 1.1 times the benchmark rate. In September 2008, the global economy suffered a major financial crisis that emanated from the US, and China was no exception. Thus, on November 17, 2008, the central government reversed the course regarding its controls on house prices and issued other measures.

With the rapid revival in China's economy, pent-up demand for houses burst into the open in the second quarter
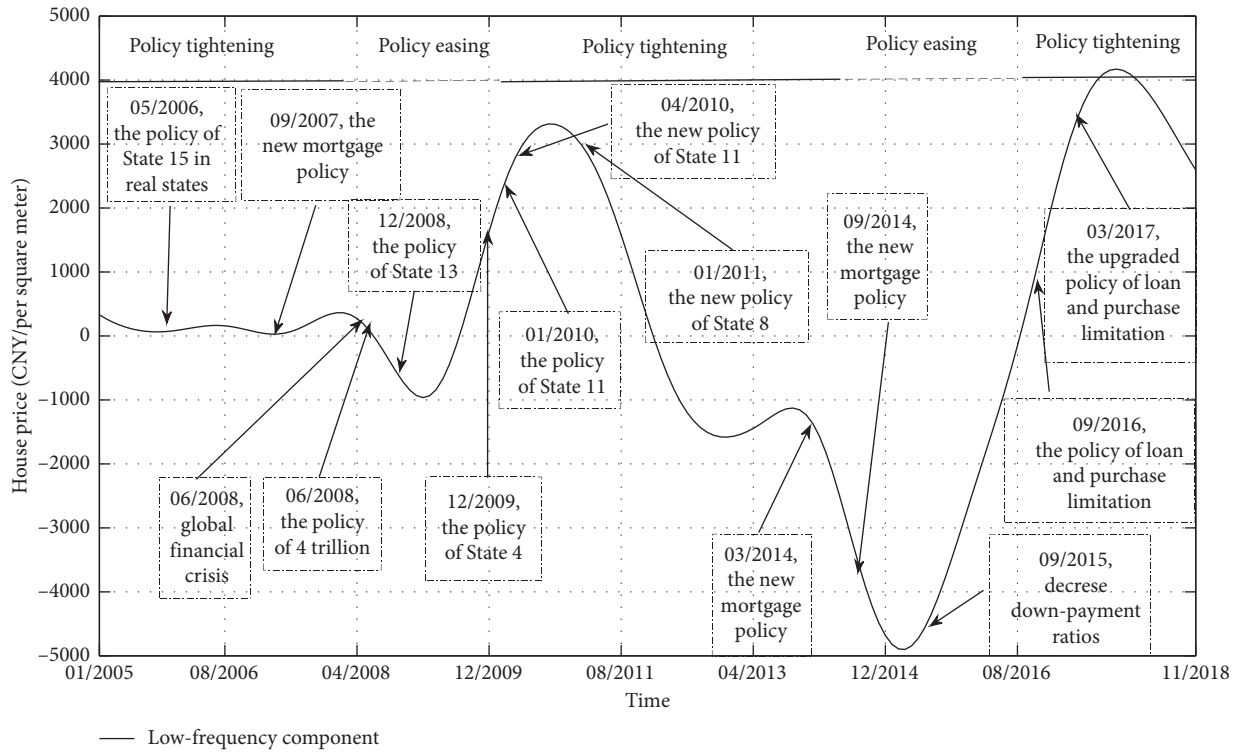
Figure 6: The low-frequency component of house prices in Beijing with some significant events.

Table 6: Real estate regulations, 2005 to 2017.

Policy direction	Date of release	Main focus of the policy
Tightening	03/26/2005	State 8: the first regulation policy, whose symbolic meaning is greater than practical meaning, in real estate.
	05/29/2006	State 15: an improved and enhanced version of State 8.
	09/27/2007	Increases down payments and interest rates for real estate loans.
Easing	12/17/2008	State 13: decreases down payments and interest rates for real estate loans.
Tightening	12/14/2009	State 4: increases supply of salable land and restrains speculation.
	01/10/2010	State 11: increases down payments and interest rates for real estate loans.
	04/17/2010	New State 10: increases the supply of salable land and strengthen indemnification houses.
	01/26/2011	New State 8: increases down payment on a second home to 60%.
	02/20/2013	New State 5: limits purchases.
Easing	09/30/2014	New mortgage policy: decreases down payments and interest rates for real estate loans.
	03/30/2015	New mortgage policy: decreases down payment on a second home to 40%.
	09/30/2015	New mortgage policy: decreases down payment on a first home to 25%.
Tightening	09/30/2016	Limits loans and purchases in 19 major cities.
	03/17/2017	Further limits on loans and purchases.

Note that the regulation policies on real estate listed here would be revised several times.

of 2009, and house prices began to trend upward again. To restrain speculation in real estate, the central and local governments responded by increasing the supply of land for sale, raising down payments, and increasing interest rates on real estate loans. This flurry of real estate regulations led to a decline in house prices in 2013. Then, in mid-2014, to achieve destocking in real estate, the government reversed course on house prices again by adjusting mortgage policies on September 30, 2014, May 30, 2015, and September 30, 2015, enabling another cycle of growth in the mortgage

industry and ballooning in house prices in 2016, with rapid, record-breaking sales and price growth. On March 31, 2016, the central and local governments introduced new real estate policies to cool demand in Tier 1 cities (currently, Tier 1 cities in China refer to the Beijing, Shanghai, Guangzhou, and Shenzhen.). And these policies on loan and purchase limitation spread to Tier 2 cities (currently, Tier 2 cities in China mainly refer to the provincial capitals in the economically developed provinces; the typical Tier 2 cities in China are Nanjing, Hangzhou, Chengdu, Wuhan, Xiamen,

etc.) on September 30, 2016. To further curb the rising house prices, enhanced and extensive regulations on loans and purchasing limits were issued on May 17, 2017.

Figure 6 shows that every big fluctuation in the low-frequency component is accompanied by one or more policies (and other events not considered in this study). For example, mortgage policies released on September 30, 2014, May 30, 2015, and September 30, 2015, respectively, and other events that occurred during this period caused sharp increases in house prices: from RMB 23,202 per sq.m. in February 2015 to RMB 38,573 per sq.m. in September 2016—an increase of RMB 15,371 per sq.m. (66.248%).

What is the exact effect of these policies and events on house prices in isolation from other factors? According to the low-frequency component, the values of this component are RMB −4,892 per sq.m. in February 2015 and RMB 245 per sq.m. in September 2016, with a lag of RMB 5,138 per sq.m. Therefore, the effect is an increase of RMB 5,138 per sq.m., less than RMB 15,371 per sq.m. compared with the original housing prices. It is conceivable that the reason for this result is that the effects of normal market disequilibrium and long-term trends, excluding mortgage policies and other important events, on house prices are an increase of RMB 10,233 per sq.m. (the difference between RMB 15,371 and RMB 5,138 per sq.m.).

4.2.4. Normal Market Disequilibrium.

In addition to long-term trends and extreme events, house prices are also influenced by short-term market fluctuation, in particular, the normal market disequilibrium between supply and demand. The effect of normal market disequilibrium is reflected in the high-frequency component, which consists of IMF1, IMF2, and IMF3. Table 3 shows that the shortest and longest mean periods of these three IMFs are 2.79 months and 15.22 months, respectively. In addition, in Table 5, the mean period of the high-frequency component is 3.08 months, indicating that the effect of normal market disequilibrium on house prices fades quickly, an average of one year. Pearson and Kendall correlation coefficients and variance contributions are quite low, 0.101, 0.062, and 2.29%, respectively, revealing that the high-frequency component is not very correlated with original house prices. This shows that the effect of normal market disequilibrium on house prices is modest.

4.3. Prediction Comparison.

The recursive method that makes use of an increasing window is used to reestimate the prediction models. To be clear, in the first round, the window size is 111, and the mentioned above procedure is executed. In the second round, the window size increases to 112, and the mentioned above procedure is executed again. After 56th round, the out-of-sample forecasts are all obtained. Even though the number of IMFs would vary with the increase in window size, the number of prediction models is always restricted to be 4 (three prediction models for single forecasting and one prediction model for ensemble forecasting) because the three composed components, instead of individual IMFs and residual, are modeled and forecasted in this study.

We attempt to explore the superiority of the proposed hybrid prediction model under the multiscale analysis paradigm to the individual SVR, ARIMA, and polynomial function used in the hybrid prediction model. Thus, the individual ARIMA, polynomial function, and SVR are selected as benchmarks. It should be noted that these three individual prediction models are performed on the original dataset instead of the decomposed dataset because the main aim of comparison is to examine the superiority of the proposed EEMD-based hybrid approach relative to the individual ones. To verify the effectiveness of the proposed EEMD-based hybrid approach, we use the housing price series in Beijing as a test sample. We use an iterative strategy to make a six-month-ahead prediction in this study.

To investigate the performance of the proposed EEMD-based hybrid approach and three competitors, we use two alternative forecast accuracy measures: the root mean squared error (RMSE) and the mean absolute percentage error (MAPE). They are defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{N}\left(x(t) - \widehat{x}(t)\right)^2}{N}}, \tag{2}$$

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x(t) - \widehat{x}(t)}{x(t)}\right| * 100\%, \tag{3}$$

where $x(t)$ is the observation at period t, $\widehat{x}(t)$ is the predicted value of $x(t)$, and N is the number of forecasting periods.

First, the original house price series is split into the estimation sample and a hold-out sample. Data from January 2005 to March 2014 are used for the estimation sample (111 observations), and 56 observations from April 2014 to November 2018 are used as a hold-out sample. Then, the model selections, particularly for SVR, are made with iterated prediction strategies in the estimation sample. Finally, the out-of-sample performance of each attained model is justified on hold-out samples. For ARIMA estimation, we develop a package forecast (the R package forecast is available at http://ftp.ctex.org/mirrors/CRAN/) using R software and use LibSVM for SVR modeling. The radial basis function (RBF) is selected as the kernel function through preliminary simulation. To determine the hyperparameters, i.e., regularization parameter C, ε in ε-insensitive loss function, and RBF kernel parameter γ, the commonly used grid search is applied here. In addition, 5-fold cross validation is used in the training process to avoid overfitting. To perform polynomial function estimation, we use the MATLAB functions polyfit and polyval.

Table 7 compares the performance of different models in all prediction horizons across two indices. The original house price and out-of-sample forecasts of four prediction models for one-, three-, and six-step-ahead forecasting are shown in Figure 7. The results lead us to make the following observations. Generally, the best model across two indices and three prediction horizons is the proposed EEMD-based

TABLE 7: The forecasting performance of the proposed approach and three alternatives.

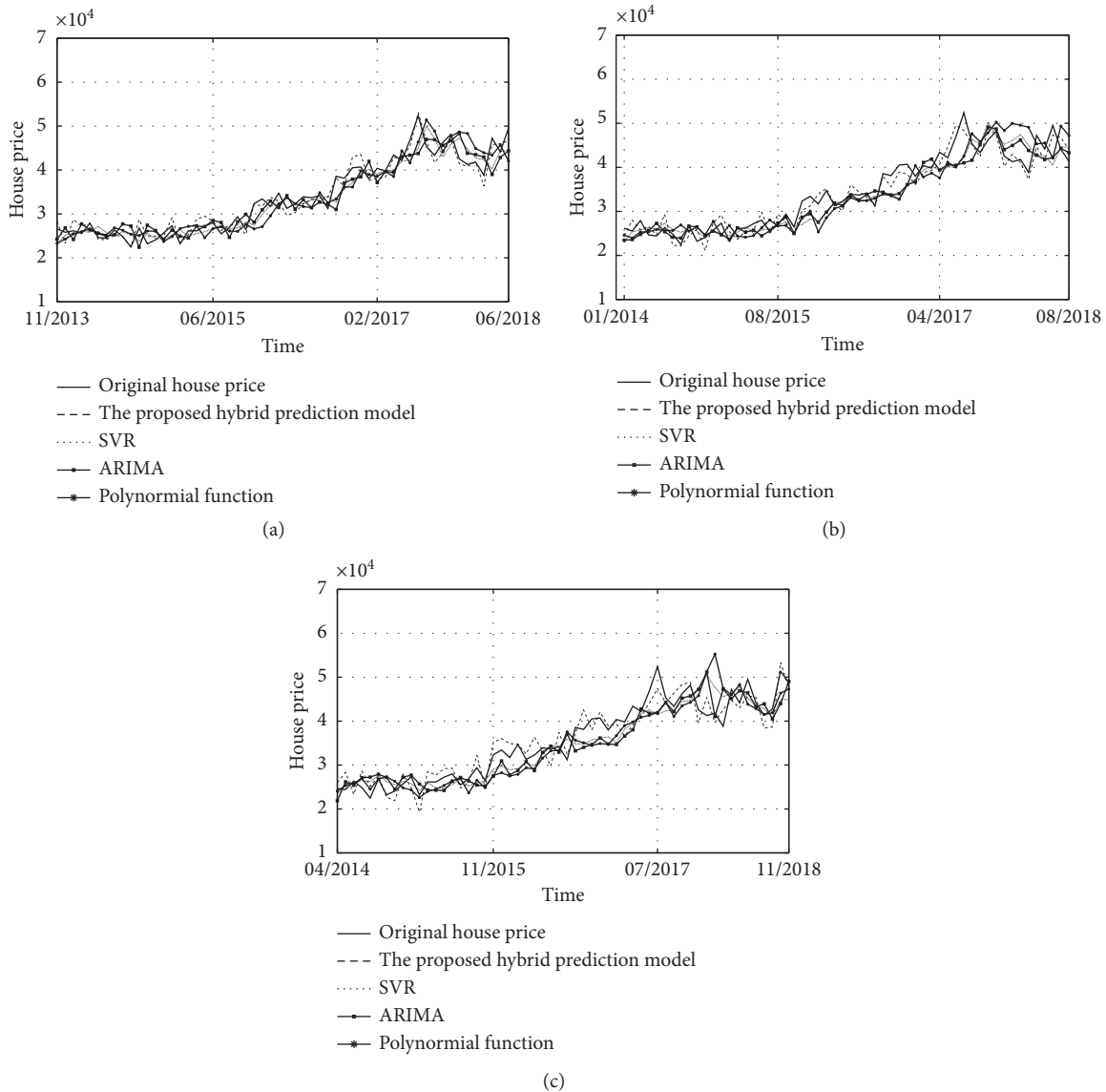| Accuracy measure	Prediction horizon	Approach			
		Proposed approach	SVR	ARIMA	Polynomial function
RMSE	One-month-ahead	3,174	4,021	4,282	4,412
	Three-month-ahead	3,524	4,592	4,834	4,715
	Six-month-ahead	4,189	4,618	4,992	4,862
MAPE	One-month-ahead	5.62%	7.18%	8.25%	8.14%
	Three-month-ahead	7.24%	10.21%	10.13%	9.21%
	Six-month-ahead	8.63%	10.15%	12.92%	10.57%

(a)

(b)

(c)

FIGURE 7: The original house price and forecasts of four prediction models on the out-of-sample period: (a) one-step-ahead forecasting; (b) three-step-ahead forecasting; (c) six-step-ahead forecasting.

hybrid approach, indicating that it outperforms the other approaches. SVR produces forecasts that are more accurate than those of the ARIMA and polynomial function. In addition, the polynomial function consistently produces the worst forecasting model.

Using the proposed EEMD-based hybrid approach, the six-month-ahead predictions of house prices in Beijing in December 2018, January 2019, February 2019, March 2019, April 2019, and May 2019 are 47,625.48, 46,298.15, 46,754.29, 47,467.85, 47,684.22, and 47,924.75 (RMB/sq. m.),

respectively. The results obtained show that the overall tendency in house prices in Beijing has been steady since the release of regulations in mid-2016.

5. Conclusions

In this study, following the "decomposition and ensemble" principle, we conduct a behavioral mechanism analysis and forecast house prices with a multiscale analysis. By integrating an ensemble empirical mode decomposition and a fine-to-coarse reconstruction algorithm, we decompose original house prices into three components: short-term fluctuations originating from the normal market disequilibrium of supply and demand; the effects of extreme events, such as the outbreak of a financial crisis and the release of regulations; and a long-term trend. Then, we examine the characteristics of these three components and develop a four-step modeling framework to forecast house prices that integrates EEMD, fine-to-coarse reconstruction algorithm, ARIMA, polynomial function, and support vector regression (SVR). The experimental results obtained in this study indicate that our proposed prediction model outperforms the alternatives, regardless of the accuracy measure and prediction horizon considered. Indeed, the RMSEs of the proposed prediction model are 3174, 3524, and 4189 for one-, three-, and six-step-ahead prediction, respectively, while the RMSEs of the SVR (the best one in the three competitors) are 4021, 4592, and 4618, respectively.

This study could be extended in many interesting directions. In particular, future studies should examine the effect of the local regulations (or the revised version) and events on house prices in a specific city and the effect of a specific regulation or event on house prices. An event analysis or difference-in-differences method could be used to explore this problem. In addition, a more sophisticated prediction strategy should be developed to forecast long-term housing prices.

Data Availability

The monthly data on the price of new residential houses in Beijing were obtained from the National Bureau Statistics (NBS) of China.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. Wen and A. C. Goodman, "Relationship between urban land price and housing price: evidence from 21 provincial capitals in China," *Habitat International*, vol. 40, pp. 9–17, 2013.

[2] B. Case, J. Clapp, R. Dubin, and M. Rodriguez, "Modeling spatial and temporal house price patterns: a comparison of four models," *The Journal of Real Estate Finance and Economics*, vol. 29, no. 2, pp. 167–191, 2004.

[3] K. E. Case and R. J. Shiller, "Forecasting prices and excess returns in the housing market," *Real Estate Economics*, vol. 18, no. 3, pp. 253–273, 1990.

[4] H. Du, Y. Ma, and Y. An, "The impact of land policy on the relation between housing and land prices: evidence from China," *The Quarterly Review of Economics and Finance*, vol. 51, no. 1, pp. 19–27, 2011.

[5] R. Gupta and A. Kabundi, "The effect of monetary policy on house price inflation: a factor augmented vector autoregression (FAVAR) approach," *Journal of Economic Studies*, vol. 37, no. 6, pp. 616–626, 2009.

[6] R. Gupta, M. Jurgilas, and A. Kabundi, "The effect of monetary policy on real house price growth in South Africa: a factor-augmented vector autoregression (FAVAR) approach," *Economic Modelling*, vol. 27, no. 1, pp. 315–323, 2010.

[7] C. Hott and P. Monnin, "Fundamental real estate prices: an empirical estimation with international data," *The Journal of Real Estate Finance and Economics*, vol. 36, no. 4, pp. 427–450, 2008.

[8] B. Hossain and E. Latif, "Determinants of housing price volatility in Canada: a dynamic analysis," *Applied Economics*, vol. 41, no. 27, pp. 3521–3531, 2009.

[9] J. M. Quigley, "Real estate prices and economic cycles," *International Real Estate Review*, vol. 2, no. 1, pp. 1–20, 1999.

[10] G.-J. Wang and C. Xie, "Correlation structure and dynamics of international real estate securities markets: a network perspective," *Physica A: Statistical Mechanics and Its Applications*, vol. 424, pp. 176–193, 2015.

[11] Z. Yang, R. R. Ren, H. Y. Liu, and H. Zhang, "Land leasing and local government behaviour in China: evidence from Beijing," *Urban Studies*, vol. 52, no. 5, pp. 841–856, 2015.

[12] L. Zhang, E. C.-M. Hui, and H. Wen, "Housing price-volume dynamics under the regulation policy: difference between Chinese coastal and inland cities," *Habitat International*, vol. 47, pp. 29–40, 2015.

[13] J. M. Clapp and C. Giaccotto, "Evaluating house price forecasts," *Journal of Real Estate Research*, vol. 24, no. 1, pp. 1–26, 2005.

[14] L. Bork and S. V. Møller, "Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection," *International Journal of Forecasting*, vol. 31, no. 1, pp. 63–78, 2015.

[15] S. C. Bourassa, E. Cantoni, and M. Hoesli, "Spatial dependence, housing submarkets, and house price prediction," *The Journal of Real Estate Finance and Economics*, vol. 35, no. 2, pp. 143–160, 2007.

[16] E. Beracha, W. M. Babajide, and K. Wang, "Forecasting residential real estate price changes from online search activity," *Journal of Real Estate Research*, vol. 35, no. 3, pp. 283–312, 2013.

[17] X. Wang, J. Wen, Y. Zhang, and Y. Wang, "Real estate price forecasting based on SVM optimized by PSO," *Optik*, vol. 125, no. 3, pp. 1439–1443, 2014.

[18] B. Park and J. K. Bae, "Using machine learning algorithms for housing price prediction: the case of fairfax county, Virginia housing data," *Expert Systems with Applications*, vol. 42, no. 6, pp. 2928–2934, 2015.

[19] M. H. Rafiei and H. Adeli, "A novel machine learning model for estimation of sale prices of real estate units," *Journal of*

Construction Engineering and Management, vol. 142, no. 2, Article ID 04015066, 2016.

[20] Y. Tan, H. Xu, and E. C. M. Hui, "Forecasting property price indices in Hong Kong based on grey models," *International Journal of Strategic Property Management*, vol. 21, no. 3, pp. 256–272, 2017.

[21] J. Chai, Y. Wang, S. Wang, and Y. Wang, "A decomposition-integration model with dynamic fuzzy reconstruction for crude oil price prediction and the implications for sustainable development," *Journal of Cleaner Production*, vol. 229, pp. 775–786, 2019.

[22] X. Zhang, K. K. Lai, and S.-Y. Wang, "A new approach for crude oil price analysis based on empirical mode decomposition," *Energy Economics*, vol. 30, no. 3, pp. 905–918, 2008.

[23] D. O. Afanasyev, E. A. Fedorova, and V. U. Popov, "Fine structure of the price-demand relationship in the electricity market: multi-scale correlation analysis," *Energy Economics*, vol. 51, pp. 215–226, 2014.

[24] G. Oladosu, "Identifying the oil price-macroeconomy relationship: an empirical mode decomposition analysis of US data," *Energy Policy*, vol. 37, no. 12, pp. 5417–5426, 2009.

[25] M. Li and Y. Huang, "Hilbert-Huang Transform based multifractal analysis of China stock market," *Physica A: Statistical Mechanics and Its Applications*, vol. 406, pp. 222–229, 2014.

[26] F. Ouyang, B. Zheng, and X. Jiang, "Intrinsic multi-scale dynamic behaviors of complex financial systems," *PLoS One*, vol. 10, no. 10, Article ID e0139420, 2015.

[27] L. Yu, J. Li, L. Tang, and S. Wang, "Linear and nonlinear Granger causality investigation between carbon market and crude oil market: a multi-scale approach," *Energy Economics*, vol. 51, pp. 300–311, 2015.

[28] B. Zhu, P. Wang, J. Chevallier, and Y. Wei, "Carbon price analysis using empirical mode decomposition," *Computational Economics*, vol. 45, no. 2, pp. 195–206, 2014.

[29] I. Antoniadou, G. Manson, W. J. Staszewski, T. Barszcz, and K. Worden, "A time-frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions," *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 188–216, 2015.

[30] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *in Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[31] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[32] L. Xian, K. He, and K. K. Lai, "Gold price analysis based on ensemble empirical model decomposition and independent component analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 454, pp. 11–23, 2016.

[33] T. Lan, "House price bubbles in China," *Research in Applied Economics*, vol. 6, no. 1, p. 86, 2014.

[34] V. Limsombunc, C. Gan, and M. Lee, "House price prediction: hedonic price model vs. artificial neural network," *American Journal of Applied Sciences*, vol. 1, no. 3, pp. 193–201, 2004.