Forecasting of the U.S. House Price Index: A Comparative Study of ARIMA, SARIMA, and ETS Models

Leyu Jia *

Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom

* Corresponding Author Email: xiaowanzi6691@gmail.com

Abstract. The U.S. housing price index (CSUSHPINSA) is an important measure of the real estate market. It reflects overall economic development, financial market activity, and changes in consumer purchasing power. This makes it valuable for research and forecasting. In this study, monthly data from January 1987 to December 2020 are used as the training set, and data from January 2021 to June 2025 are used as the testing set. Time series models are built to forecast the U.S. housing price index. Three models are applied: ARIMA, SARIMA, and ETS. They are evaluated through parameter estimation, model fitting, and forecasting comparison. Their performance is assessed in terms of trend description, interval coverage, and forecasting accuracy. Results show that all three models continue the long-term upward trend of the housing price index. ARIMA underestimates growth during the testing period, while ETS gives wide intervals and means forecasts that deviate from the actual values. In contrast, SARIMA produces forecasts that match the observed trend closely. Its error measures, including RMSE, MAE, and MAPE, are much lower than those of the other models. In conclusion, the SARIMA model not only reflects the long-term trend but also captures seasonal fluctuations effectively. It is the most suitable forecasting method in this study and provides useful reference value for housing market monitoring and policy making.

Keywords: Prediction; ARIMA; SARIMA; ETS; CSUSHPINSA.

1. Introduction

The real estate market has important economic and social significance worldwide. At the economic level, real estate is not only a key part of the national economy, but its investment and construction activities can also directly drive fixed asset investment and employment growth [1]. In addition, housing price fluctuations are closely linked to macroeconomic performance. They influence household consumption and financing conditions through the wealth effect and the balance sheet channel [2]. At the financial level, real estate is closely connected to the monetary and credit system. Rising housing prices increase collateral values and support credit expansion. Falling prices, however, may weaken bank asset quality, trigger credit contraction, and accumulate risks. In this way, the real estate market can become a potential source of financial instability [3]. At the social level, real estate is the main form of household wealth. Compared with financial wealth, housing wealth has a stronger marginal effect on consumption. Therefore, housing price fluctuations affect household consumption and wealth distribution. They may also influence social stability through changes in expectations and psychological channels [4]. In summary, the fluctuations of the real estate market are not only important for the stability and growth of the economy, but also strongly shape social welfare and long-term development prospects.

After clarifying the importance of the real estate market, research has extended to the economic, financial, and social dimensions. In recent years, many studies have provided new findings and challenges. At the economic level, scholars note that housing market volatility has a major impact on the macroeconomy. Liquidity frictions in housing transactions are one of the main mechanisms behind price fluctuations [5]. The housing market is also closely linked to the business cycle, reflecting macroeconomic conditions [6]. At the financial level, research highlights the link between real estate and financial risks. The growth of housing finance and credit expansion are key drivers of real estate cycles and may lead to systemic risks [7]. Studies further show two-way feedback between housing prices, money, and credit, where market instability spreads to the financial system through collateral and wealth effects [8]. At the social and policy level, housing affordability has received rising attention. It directly affects household welfare and social equity [9]. Moreover, it is closely linked to economic growth, with policy interventions serving as a key factor [10]. The COVID-19 shock has added new perspectives, showing that the housing market is vulnerable to unexpected risks [11]. Overall, although progress has been made in economic, financial, and social research, there is still room to improve cross-domain forecasting of housing market fluctuations and related policy responses.

This study focuses on forecasting housing market prices and selects the U.S. national housing price index (CSUSHPINSA) as the research object. In this paper, the ARIMA model, the SARIMA model, and the ETS model are built. These three models are compared with their forecasting performance. Also, parameter estimation and error indicators are applied. By measuring them, how well these models capture long-term trends, prediction accuracy, and interval coverage are assessed. The findings suggest that SARIMA is the best model. Moreover, the most reliable result is achieved because it is able to reflect both the seasonal fluctuations and the upward movement of the housing price index. Overall, the dynamics of the housing market are provided by this research, and future academic work and policy considerations are served with useful references.

2. Data

Widely, the CSUSHPINSA index is used to study the real estate market. Compared with a single city or region, it is different from them, which reflects price movements across major areas of the United States. Thus, based on a larger sample, the overall trend in the national housing market is offered. Significantly, it is also possible to trace patterns in the country's economic development by this index. Generally, changes in housing prices are not only reflected in household purchasing power but also connected with financial activity and monetary policy. For this reason, the index has strong predictive value and significance in research.

The CSUSHPINSA data used in this study comes from FRED. To better analyze the long-term trend of the U.S. housing market, the dataset covers monthly observations from January 1987 to June 2025. December 2020 is chosen as the dividing point to separate the training set (January 1987 to December 2020) and the testing set (January 2021 to June 2025). The CSUSHPINSA index during the full sample period is shown on Fig. 1.

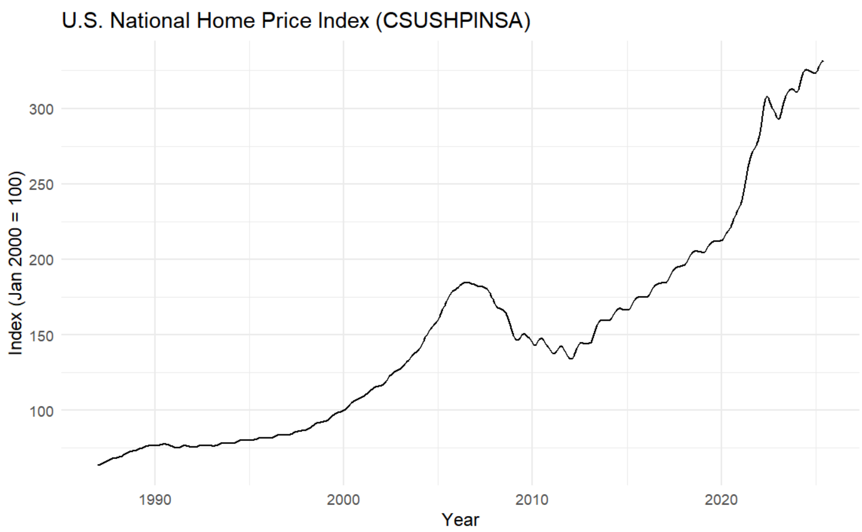

Fig. 1 CSUSHPINSA

Fig. 1 shows the trend of the CSUSHPINSA data. Overall, the index follows a steady upward trend. During the 2008 financial crisis, however, it experienced a clear decline. After that, the upward movement gradually resumed. From 2020 to 2022, the growth rate of the housing price index rose sharply, and it reached its highest point in June 2025. The basic descriptive statistics of the sample period are presented in Table 1.

Table 1. Descriptive Statistics

series	mean	SD	max	min	kurtosis	skewness
CSUSHPINSA	150.1	70.65	331.52	63.73	3.16	0.92

According to Table 1, the maximum value of the data is 331.52, the minimum value is 63.73, the mean is 150.1, and the standard deviation is 70.65. This shows that there are large fluctuations in the sample values. The kurtosis is 3.16, which is greater than 0, indicating that compared with a normal distribution, the data distribution has a more peaked shape, meaning that the fluctuations are concentrated around the mean. The skewness is 0.92, which is greater than 0, showing that the distribution is skewed to the right.

3. Method

3.1. ARIMA Model

The ARIMA model is a classical method for time series forecasting. The model is relatively simple and only requires endogenous variables. This feature allows it to capture the dynamic patterns in the data even when external explanatory variables are absent. In particular, for non-stationary data with a certain trend, the ARIMA model predicts future values by analyzing autocorrelation and applying differencing to past data.

First, the differencing part (the I component) is used to remove trends and seasonal effects so that the series becomes stationary. Second, both the autoregressive (AR) term and the moving average (MA) term are incorporated by this model. They are used to fit the data. What is more, these components describe how current values are related to previous observations as well as past forecast errors. In this way, the model can be applied to generate predictions for future periods. The specific formula is as follows:

$$\phi_p(L)(1-L)^d y_t = \theta_q(L)\varepsilon_t \tag{1}$$

In this study, y_t represents the observed value of CSUSHPINSA for month t. Also, it covers the period from January 1987 to June 2025. The parameter d denotes the differencing order, and here a first-order difference is applied to achieve stationarity. The term "p" refers to the autoregressive order. It indicates the relationship with the current housing price index and its previous values. The parameter q stands for the moving average order. What's more, it captures the influence of past forecast errors on the present value. Finally, ε_t means the random shocks or unexplained variations in the housing price index.

3.2. SARIMA Model

Compared with the ARIMA model, the SARIMA model is able to account for more complex patterns for seasonal time series data. Accurately, apart from general trends and random movements, it can also describe the cyclical patterns of the data clearly. Furthermore, the U.S. housing price index (CSUSHPINSA) illustrates this absolutely: while it reflects a long-term upward trend shaped by macroeconomic conditions, it is also influenced by market cycles, seasonal shifts in housing demand, policy changes and so on. As a result, the ARIMA model cannot significantly capture these seasonal features than the SARIMA model.

Based on the ARIMA model, the SARIMA model adds more seasonal elements, such as seasonal autoregression, seasonal differencing, and seasonal moving averages. With these additions, it can capture the autocorrelation structure of cyclical data more effectively. The specific formula is as follows:

$$\Phi_P(L^s)\phi_p(L)(1-L)^d(1-L^s)^D y_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t \tag{2}$$

In this study, y_t is used to denote the observed value of CSUSHPINSA in month t. The parameters p, d, and q describe the non-seasonal part of the model: p refers to the first-order autoregressive term, d is the first-order differencing term, and q is the second order moving average term. Also, the seasonal part is expressed by P, D, and Q. P corresponds to a zero-order autoregressive term, D is a first-order seasonal difference, and Q means a second-order seasonal moving average. The symbol s indicates the monthly frequency, which is set to a 12-month cycle. ε_t represents the part of the housing price index that cannot be explained by the model.

3.3. ETS Model

The ETS model is a forecast approach. It is based on exponential smoothing. The principle of this model is that the more recent observations can contain more information about the future. The result should be given higher weights. Because of this property, the model is usually applied to series that display trends, seasonality, and random fluctuations. Through its smoothing parameters, ETS declines the weights of past values in an exponential form. This allows recent data to play a larger role in prediction and makes the model more responsive to new developments.

There are three key advantages that make the ETS model useful. First, it uses three main components: Error, Trend, and Seasonal, to describe short-term disturbances, long-term movements, and cyclical variations. Second, compared with ARIMA-type models that rely on differencing and autoregressive structures, the ETS model is more straightforward in design, requires fewer parameters, and is easier to understand, which allows for faster forecasting and modeling. In summary, the model is more flexible in capturing short-term fluctuations and handling economic and financial data with both trends and seasonality. Therefore, for time series such as the housing market, which shows an overall upward trend with cyclical fluctuations, the ETS model has high applicability. The specific formula is as follows:

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t \tag{3}$$

In this model, y_t still refers to the observed value of CSUSHPINSA in month t. l_{t-1} is the level component, representing the long-term average level of the housing price index in the previous month. b_{t-1} is the trend component, indicating the direction and speed of changes in the housing price index in the previous month. s_{t-m} is the seasonal component, with m = 12, representing the cyclical fluctuations of the housing market within a year. ε_t represents the fluctuations in the housing price index that are not explained by the model.

4. Results

In this study, the data from January 1987 to December 2020 are used as the training set, and the data from January 2021 to June 2025 are used as the testing set. The parameters obtained by fitting the training data are presented in Table 2.

Table 2. Parameter Fitting Results

Model	Parameter1	Parameter2	Parameter3
ARIMA	3	1	2
SARIMA	1	1	2
ETS	M	Ad	N

Based on these parameter settings, this study applies three models: ARIMA (3,1,2), SARIMA (1,1,2) (0,1,2) (12), and ETS (M, Ad, N), to forecast the U.S. housing price index for the testing period from January 2021 to June 2025. The forecasting results are shown in Fig. 2, where the predicted trajectories of the three models are displayed together with the actual observations.

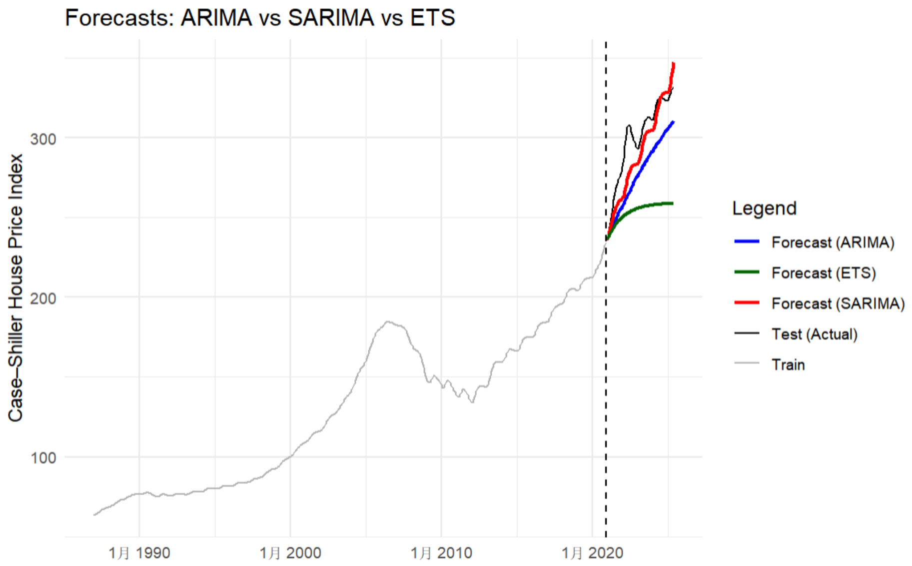

Fig. 2 Predicted vs Actual Values

From the perspective of forecast trends, all three models continue the long-term upward movement observed in the training set (January 1987 to December 2020). However, there are significant differences in their specific forecast performance. For the ARIMA and ETS models, both show clear shortcomings compared with the SARIMA model in terms of accuracy and interval coverage. The ARIMA model not only produces forecast means that deviate from the actual trend, but also continuously underestimates the growth during the testing period, resulting in systematic bias between the central forecasts and the real values. The ETS model performs even worse. Its prediction intervals are too wide, and its mean forecasts deviate greatly from the actual trend, making it difficult to provide reliable reference information. By contrast, the SARIMA model performs the best in forecasting. Its forecast has almost complete overlap with the actual observations, accurately following the fluctuations of the housing price index. It captures rapid growth during the testing period with high sensitivity, and its forecast intervals reasonably reflect future uncertainty, showing the best fit and predictive performance.

To make a quantitative comparison of forecasting performance, error metrics are applied. The SARIMA model shows clear advantages, as its RMSE (13.61), MAE (11.03), and MAPE (3.70%) are all significantly lower than those of ARIMA and ETS. This further confirms its superiority in forecasting accuracy, as shown in Table 3.

Table 3. Comparison of Forecasting Errors

Model	RMSE	MAE	MAPE
SARIMA	13.61	11.03	3.70%
ETS	50.13	46.25	14.95%
ARIMA	55.45	52.04	16.90%

Overall, considering both qualitative and quantitative analysis, the SARIMA model demonstrates the best forecasting ability. Compared with ARIMA and ETS, the forecast curve of SARIMA fits the actual trend most closely. At the same time, the model also performs much better on error metrics such as RMSE, MAE, and MAPE.

In fact, the results indicate that the SARIMA model captures the long-term trend of the U.S. housing price index. Also, it reflects seasonal fluctuations effectively. Therefore, it is the most suitable forecasting method in this study.

5. Conclusion

This study uses the U.S. housing price index from January 1987 to December 2020 as the training set. Data from January 2021 to June 2025 are used as the testing set. Three time series models are compared: ARIMA, SARIMA, and ETS. The results show that all three models capture the long-term upward trend of the housing price index. However, their forecasting accuracy is quite different. The ARIMA model underestimates the growth during the testing period. The ETS model produces wide intervals and means forecasts that deviate from the actual trend. In contrast, the SARIMA model matches the observed values closely. It also performs best in terms of error measures such as RMSE, MAE, and MAPE. Overall, the results show that SARIMA is the most suitable model in this study, as it reflects the long-term trend and captures seasonal changes accurately.

Although the SARIMA model shows the best performance among these three models, this study is not without limitations. Only the internal structure of the time series is considered. External factors such as interest rates, inflation, or employment are not included. This may reduce the explanatory and predictive power of the models. Future studies could add such external variables to build a more complete forecasting framework and improve both stability and practical value.

References

[1] Case K E, Glaeser E L, Parker J A. Real estate and the macroeconomy. Brookings Papers on Economic Activity, 2000(2): 119-162.

[2] Case K E, Quigley J M, Shiller R J. Comparing wealth effects: the stock market versus the housing market. Topics in Macroeconomics, 2005, 5(1).

[3] Goodhart C, Hofmann B. House prices, money, credit, and the macroeconomy. Oxford review of economic policy, 2008, 24(1): 180-205.

[4] Benjamin J, Chinloy P, Jud D. Why do households concentrate their wealth in housing. Journal of Real Estate Research, 2004, 26(4): 329-344.

[5] Ngai L R, Sheedy K D. The Ins and Outs of Selling Houses: Understanding Housing-Market Volatility. International Economic Review, 2024, 65(3): 1415-1440.

[6] Kohler K, Tippet B, Stockhammer E. House price cycles, housing systems, and growth models. European Journal of Economics and Economic Policies, 2023, 20(3): 461-490.

[7] Ryan-Collins J. Breaking the housing–finance cycle: Macroeconomic policy reforms for more affordable homes. Environment and planning A: economy and space, 2021, 53(3): 480-502.

[8] Agnello L, Castro V, Sousa R M. The housing cycle: what role for mortgage market development and housing finance. The Journal of Real Estate Finance and Economics, 2020, 61(4): 607-670.

[9] Haffner M E, Hulse K. A fresh look at contemporary perspectives on urban housing affordability. International Journal of Urban Sciences, 2021, 25(sup1): 59-79.

[10] Anthony J. Housing affordability and economic growth. Housing policy debate, 2023, 33(5): 1187-1205.

[11] Mohammed J K, Aliyu A A, Dzukogi U A, et al. The impact of COVID-19 on housing market: A review of emerging literature. International Journal of Real Estate Studies, 2021, 15(2): 66-74.