

Chinese Housing Prices Prediction using Autoregressive Integrated Moving Average Model

Hongyu Wan *

Hwa Chong International School, 269783, Singapore

* Corresponding Author Email: wan.hy@stu.hcis.edu.sg

Abstract. This paper investigates the application of the Autoregressive Integrated Moving Average (ARIMA) model to predict future trends in Chinese housing prices. The Chinese real estate market, characterized by its volatility, especially during the post-COVID-19 period, presents a complex environment for buyers and investors. The paper investigates how the ARIMA model is employed to make informed predictions in this uncertain market. Although it has some limitations, such as a heavy reliance on historical data and insensitivity to unexpected macroeconomic shifts, the ARIMA model offers a structure for understanding and anticipating housing price trends. The paper integrates various data sources, including long-term housing price statistics, to build a comprehensive ARIMA model tailored to the nuances of China's housing market. This essay demonstrates the ARIMA model's utility in aiding stakeholders to make more confident and informed decisions in an increasingly complex and unpredictable market. The analysis further suggests refinements to the ARIMA model, considering the multifaceted nature of the housing market, influenced by macroeconomic factors and public perception. The final goal is to enhance the model's accuracy and reliability, making it an indispensable tool in economic and policy decision-making in China's evolving real estate landscape.

Keywords: Housing price; ARIMA model; Chinese.

1. Introduction

China's real estate market has experienced volatility after the end of the COVID-19 pandemic. In August 2023, market house prices fell slightly by 0.29% [1]. Although the decline was modest, it affected the confidence of potential buyers. Due to the declining price action in the market, many people are postponing their buying decisions because they want to wait for the lowest point before making a purchase [2]. In this situation, informed and careful decision-making is essential, and predictive models such as the Autoregressive Integrated Moving Average (ARIMA) model have become a helpful tool for analysts and consumers. The real estate market is affected by many factors, according to government policies in the macroeconomy and local house supply and demand dynamics in the microeconomy [3].

As people need to put a significant financial outlay involved in purchasing a house, this is a long-term investment, so the potential homeowners are very cautious. They seek the best investment timing and selection to ensure maximum returns. The prediction model will predict the general direction of changes in housing prices in the complex housing market [4]. The autoregressive integrated moving average (ARIMA) model is a statistical analysis model that uses time series data to understand the data set better or predict future trends [5].

This model includes three parts: The first part is Autoregressive (AR), the model's feature that exploits the premise that past values are related to current values. It will assume that previous house prices can provide clues to future prices. The second part is integration (I); this component of ARIMA involves differentiating the data a certain number of times to achieve stationarity, i.e., the statistical properties of a time series do not depend on the state of the time at which the series is observed. The third part is the moving average (MA); this aspect focuses on the relationship between observations and the residuals of a moving average model applied to lag observations [6]. It helps remove the noise of randomness in time series data. It also has many limitations. For example, the autoregressive model has many limitations: it uses data to make predictions. Time series data must be stationary.

Autoregression is only suitable for predicting phenomena related to its previous period (autocorrelation of time series) [7].

The application of ARIMA models to forecasting house prices faces some challenges. The model relies on historical data. In a post-pandemic world where market behavior may have structurally changed, past data may only sometimes be a reliable and effective indicator of future trends. Furthermore, the model itself does not account for unexpected macroeconomic government policy interventions like indirect taxation that could significantly impact the housing market [8]. Although there are some challenges, the ARIMA model remains a powerful tool, especially when combined with other models and market analysis. It enables potential buyers to make informed decisions based on modal instead of guesswork [9].

In conclusion, as China's real estate market enters its post-pandemic recovery phase, forecasting models such as ARIMA will be critical to individuals and institutions in the real estate sector. Provides forecasts for the future based on a thorough analysis of historical trends. ARIMA modal will help stakeholders make more confident, stable, and reliable decisions in the property market. Like all models, it is not a crystal ball but a robust and essential guide to buying houses in a more complex and unpredictable market [10].

2. Methods

2.1. Data Source

The data for this literature is collected from the Fred website to select the unemployment rate data from April 2005 to January 2023 with 71 observations.

2.2. Variable Selection

Since China's housing price data is a long-term statistical data, a long-term period of data was selected for research and analysis. Find out the main patterns of these trend changes through time series research, as shown in Figure 1 below.



Fig 1. Housing price in China.

Some summaries of changes in housing prices in China. First, according to the trend line in Figure 1, we can see that China's housing prices are on an upward trend. Second, according to the data points, the data fluctuates wildly.

2.3. Data Processing

In the ACF graph (Figure 2), most of the stems in the first few lags are above the zero line, indicating a positive correlation at these lags. As the time gap decreases, the correlation decreases. This suggests that past values influence future values, but this influence diminishes as the time gap increases.

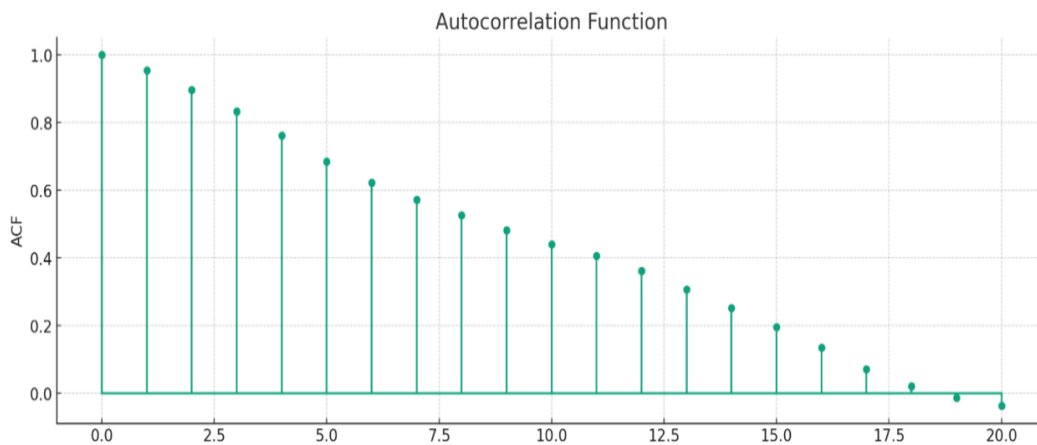


Fig 2. ACF plot.

2.4. Method Selection

The ARIMA model was chosen for its ability to handle non-stationary time series data, typical of real estate markets, including China, where prices show upward trends and cycles over time. It also achieves stationarity by differentiating the data and then captures dynamics through autoregression (relating each observation to its previous value) and moving averages (removing noise from random fluctuations). Its strength lies in its simplicity, making it a powerful and convenient forecasting tool that can be adapted to the unique characteristics of China's housing price data and does not require complex transformations. Specifically, ARIMA's ability to model time dependencies and shocks for volatile real estate markets through its integrated structure enables it to make informed forecasts. This applicability to economic and financial time series, coupled with the model's customizability and efficiency of fitting and validation, makes ARIMA a favorable choice for forecasting the trajectory of housing prices in China.

3. Results and Discussion

3.1. Forecasting Results

Figure 3, generated from the ARIMA (1, 1, and 1) model, forecasts the future value of China's housing price index for 2023 and 2024. Figure 3 represents the historical data points of China's housing price index by a line graph, showing the trend up to January 2023. Then, a forecasted trend is extended from the last historical data point (2023.1.1) into 2025.1.1, with the predicted values plotted as a separate line with a distinct style (dashed and colored red).

Table 1 provides the value of data points shown in Figure 3. It lists the predicted index values for each quarter, depending on the expected trend according to the ARIMA model. The forecasted values suggest a stop of uptrend and a slight decrease over the period, which are reflected in the downward slope of the forecast line on the chart.



Fig 3. Predicts the value of the housing price.

Table 1. ARIMA Model Results.

Time	Point forecast	Lo 95	Hi 95	Lo 80	Hi 80
2023-04	104.0535	101.3708	106.7361	102.2994	105.8075
2023-07	103.9388	99.473	108.4046	101.0188	106.8589
2023-10	103.8913	97.9372	109.8455	99.9981	107.7845
2024-01	103.8717	96.6478	111.0955	99.1483	108.5951
2024-04	103.8635	95.5307	112.1963	98.415	109.312
2024-07	103.8601	94.5378	113.1825	97.7646	109.9557
2024-10	103.8587	93.6378	114.0796	97.1756	110.5418
2025-01	103.8582	92.8098	114.9065	96.634	111.0823

3.2. Check Residuals

The ACF chart (figure 4) shows a strong initial correlation that gradually decreases as the number of lags increases. However, the ACF value remains positive and significant over multiple lag periods. This indicates a strong, persistent trend or seasonality in the underlying data that is not fully explained by the model, requiring additional differencing of the model to achieve stationarity.

The PACF chart (figure 5) shows a significant peak at the first lag, while subsequent lags fall within the desired confidence interval. The significant peak in the PACF chart at the first lag period suggests that it may be appropriate to include a first-order autoregressive term in the model, as it indicates that the value at a one-time point is predictive of the next time point after accounting for intermediate values. Value.

Combining the observations from both graphs, it is possible that the model may become better by adding additional differences, reducing trend and seasonality effects (as shown in the ACF graph), while also adding an autoregressive term (as shown in the PACF graph). Additionally, the lack of clear cutoff points in the ACF chart may indicate that adding a moving average component may also be beneficial.

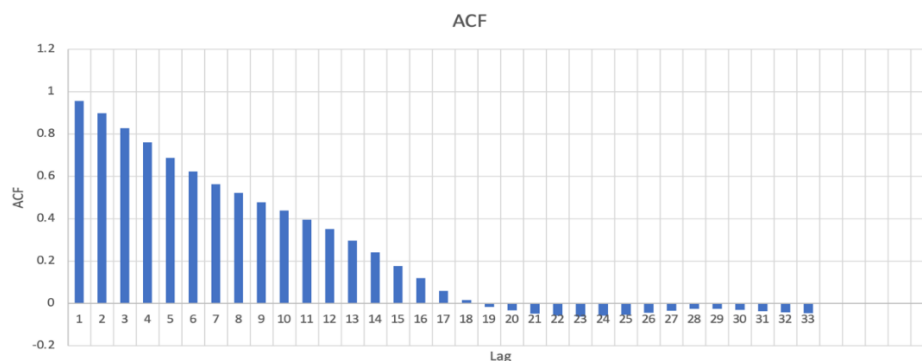


Fig 4. ACF graph of residuals.

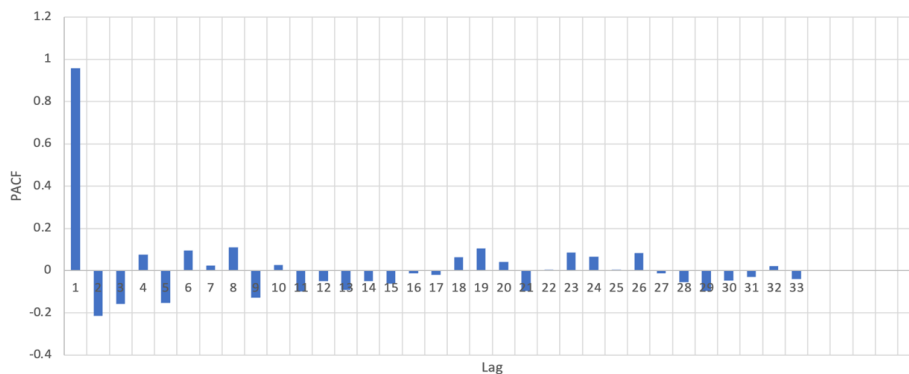


Fig 5. PACF graph of residuals.

3.3. Critical Thinking

While the ACF figure shows a decreasing yet persistent correlation over multiple lag periods, it suggests that the model may not fully capture the complexity of the market's dynamics. The PACF figure's significant peak at the first lag points to the potential benefit of incorporating an AR (1) component. However, the significant autocorrelation in the ACF figure shows that the model may benefit from the addition of further lags in the AR or MA components or from a more nuanced differencing strategy to model the data's trend and seasonality better.

4. Conclusion

The figure plots from the ARIMA model show the areas for improvement when it comes to modelling China's housing prices. The detectable autocorrelation within the residuals suggests that the model's current iteration might be too rudimentary, failing to encapsulate critical elements of market behavior that dictate housing prices. This shows an oversimplification leading to a misinterpretation of complex economic realities. As such, it becomes imperative to refine the ARIMA model to better mirror the multifaceted nature of the housing market.

To fix the non-stationarity, further differencing could be considered. Non-stationarity in time series data often leads to spurious relationships, which can cause wrong model outcomes. So, by using additional levels of differencing, the model can better account for underlying trends or cyclic patterns characteristic of economic data. Moreover, more autoregressive (AR) or moving average (MA) terms might provide stronger support for viewing the market's temporal dependencies. This would help capture the inertia or momentum often observed in housing prices, where past values have a prolonged impact on future prices.

Incorporating external variables could lead to a more holistic model. Housing markets do not operate in isolation; they are influenced by many macroeconomic factors, such as GDP growth rates, interest rates set by central banks, consumer confidence in the housing market, government housing policies, and population movement. All these factors affect the money spent on the house by individuals, ultimately affecting housing demand and prices. For example, a central bank's interest rate policy can significantly affect mortgage rates, which in turn influences the affordability of houses to people and the demand for housing. Similarly, changes in government housing policies, such as subsidies for first-time homebuyers or restrictions on foreign investments in real estate, can significantly alter market dynamics.

Moreover, the social perspective includes the public's perception of real estate as an investment. As housing prices show an upward trend until 2023, more people may use real estate as an investment to earn profit. This societal trend can shift annually, contributing to the volatility observed in the ACF plots.

In conclusion, the way to refine the ARIMA model is twofold: it involves improving the model's internal factors like autoregressive (AR), moving average (MA) and seasonal components through

statistical adjustments and broadening its scope to include external, influential factors like effects in macroeconomic area. Only then can the model serve as a true barometer of the housing market, providing stakeholders with the insights necessary to make informed decisions. The aim is to construct a model that fits the past data well and anticipates future trends with a reasonable degree of accuracy, thus becoming an indispensable tool in the arsenal of economists, policymakers, and investors.

References

- [1] Lianhe Zaobao. China's housing prices are falling at a faster pace; analysts believe that loosening policies cannot completely rescue the market. Working paper, 2023.
- [2] Elmassah S, Bacheer S, Hassanein E. US consumers' confidence and responses to COVID-19 shock. *Review of Economics and Political Science*, 2023, 8: 186-207.
- [3] Li Xinchu. Prediction and Analysis of Housing Price Based on the Generalized Linear Regression Model. Working paper, 2022.
- [4] Andrea L, Schaffer, Timothy A. Dobbins, and Sallie-Anne Pearson "Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Medical Research Methodology*, 2021, 21.
- [5] Fan Jianshuang, Zhou Lin, Yu Xiaofen. The impact of land finance and land market development on urban housing prices. *Geographic Science*, 2021, 41(5): 9.
- [6] Li Hengke, Yuan Yuan, Liu Jing, Zheng Yueyan. Analysis of housing price prediction in Wuhan based on multiple linear regression model. *Science and Mass*, 2021, 1: 279-280.
- [7] He Zenghua, Sui Shumin, Wang Jinglei. The Impact of Housing Prices on Enterprise Innovation: Based on the Perspective of Innovation Investment and Talent Policy. *Scientific Research*, 2022, 40(12): 13.
- [8] Chen Yao, Chen Xiangman. Empirical analysis of the impact and spatial effects of housing prices and the ratio of housing prices to income on China's urbanization. *Economic Geography*, 2021, 4: 9.
- [9] Zhao Wei, Wu Donglin, Zhao Minjuan. Heterogeneity expectations, individual interactions, and housing price fluctuations. *Financial Forum*, 2022, 2 (10): 10.
- [10] Li Huan, Dong Na, Pan Min, Yu Rui, Xiong Feng. Prediction of housing price index based on PCA-ARDL-BP neural network. *Construction Economics*, 2022, 43(1): 759-763.