Housing Supply under Long-term Uncertain Situation: A Counterfactual Framework Based on ARMA Model

Yuting Liu[1, *, †], Enming Zhang [2, †]

[1] Department of Mathematics and Statistics, Fuzhou University, Fuzhou, 350108, China

[2] Department of Mechanical and Engineering, Xiangtan University, Xiangtan, 411105, China

[†]These authors contributed equally.

*Corresponding author: 031901327@fzu.edu.cn

Abstract. The monthly supply of houses is a critical part in housing markets in US. It reflects the condition of economy and conversely exerts an influence to the economy. It is an important number for sellers to have in mind when planning the sale of a property investment and buyer can based on it to make their decision in making purchases. Across the United States, the COVID-19 pandemic has resulted in considerable confirmed cases every month which might influence people's decisions in changing locations and buying new houses, thus affected the variation of the monthly supply. This paper judges the negative impact of COVID-19 on the real estate industry in the United States by establishing the difference in DF value obtained by ARMA model, discusses the possible reasons for the fluctuation of new house supply in the United States, and puts forward some suggestions on how to deal with the crisis of the real estate industry in similar situations in the future, so as to avoid the economic downturn of real estate enterprises in similar situations and achieve the stability of the housing economy. Data analysis process was conducted by the use of R statistical programming.

Keywords: Monthly supply, new houses, time series, ARMA, COVID-19, R.

1. Introduction

The real estate industry is the basic carrier of the national economy, which directly or indirectly drives many industries and constitutes an important part of the wealth of the whole society. In the U.S. real estate market, the epidemic has made it difficult for those who have houses to list for sale, resulting in a small number of houses on the market. The epidemic also caused a shortage of construction materials and labors, and the supply of new houses could not keep up with the demand. Besides providing a barometer of local market housing demand and supply balance, monthly supply of homes can provide the basis for calculating the home sales rate or home absorption rate. This is a very useful metric for analyzing a residential property or a project for investment purposes [1]. The frantic pace of activity seen over the previous two years of the pandemic has left housing markets in need of a reset. While new houses during earlier COVID days became extremely appealing due to their improved usefulness and technology, as well as the severe lack of available inventory. Transactions have been slowed down by high pricing, but an improvement in the supply pipeline could ease the burden on consumers' wallets [2]. This paper aims to figure out if the confirmed cases of pandemic had a significant impact on the variation of monthly supply of new houses. Some articles have argued impacts on housing prices, private rental housing prices and the whole development of American economy. Another article on American real estate price index shows that American homeowners are reluctant to sell their houses by analyzing the record equity level, which indirectly leads to the urgent demand for new houses in the United States.

To determine the effect of the COVID-19 on the monthly supply of new houses in the US from January 2020 to June 2022. The analysis was conducted under a counterfactual framework. Assuming there was no pandemic, the ARMA model was created using data prior to the pandemic, and it was then used to forecast variations of the monthly supply during the pandemic time. The trend between the actual value and the fitted value is the same, therefore discrepancies between them can be interpreted as the pandemic's effects on the monthly supply. The pandemic's impact is then shown when the ARIMAX model includes the monthly increase in confirmed cases from January 2020 to

June 2022 as an exogenous variable and produces a new fitted value. Next, the student T-test was applied to evaluate disparities between the actual value and two fitted values, from which it is intuitively clear whether COVID-19 has a significant impact. By establishing a numerical model, this paper judges the situation of American real estate under the influence of the COVID-19 epidemic, predicts and explores the influencing factors, and finally puts forward suggestions for the future of the American real estate industry.

Herein, the following sections of this paper proceeds follows: Part 2 introduces methodologies used in this paper, including intuitive analysis of data, unit root test and model specification. Part 3 is the display of empirical results and its analysis. Some limitations of this research are listed in the discussion section and some suggestions are provided for consumers, investors, and policies in the conclusion.

2. Methodologies

2.1.Intuitive Analysis of Data

We abstracted the time series of monthly supply of new houses in US from American Federal platform [3], chose data from January 2000 to June 2022 with the range of 22 years. It should be emphasized that the numbers represent national averages. It is fair to assume that there are variations of this metric around these numbers in the different metropolitan housing markets of the country. However, as there are no metropolitan-specific data on this ratio we cannot provide any insights to the extent of these variations. From Figure 1, the monthly supply of homes started rising steadily from the second quarter of 2005, indicates the national housing market started to become softer. It reached its peak of 11.2 by the fourth quarter of 2008. From the first quarter of 2009, months supply of homes embarked on a steady downward path which bottomed out in the second quarter of 2013. The remarkable about this bottoming out is that it occurred at 4.1 months, the same number from which this big cyclical swing started in 2005 Q2. After the second quarter of 2013 it has been fluctuating between 5 and 6 months with a spike above 6 months in 2018. As of the third quarter of 2019, the last data point, it was down to 5.3 months [4].

The plot of the time series shows some sudden changes (Figure 1), particularly the crescendo in 2008/2009, and a strong seasonality of one year, it reaches a relatively high magnitude in the middle of the year, perhaps at the beginning of a year, some new houses had hung for a long time, sellers provided a discount and buyers had larger negotiation space, the pace of the selling of new houses increased, so the monthly supply was low. However, in the process of getting near to the mid of the year, there was an accretion of new houses in housing market, and sellers used to have high expectations which led to exorbitant prices of new houses, so the monthly supply reached the highest level in a year. Afterwards, for the tax rebate season is in April in the United States, which usually improves the purchasing power of buyers, which was manifested in the decreasing of monthly supply after the mid of a year. After the outbreak of COVID-19 at the beginning of 2020, with the less obvious seasonality, the monthly supply first reached a low point, then appeared an upward tendency till now. As the chief economist Daryl Fairweather said: “The demand for home purchase is supported by many factors, including the fact that as the pandemic started, the price of renting a house increased a lot, which led some first-time buyers to buy a new house under this situation.” [5] However, with the development of pandemic, there was an continuous shortage in the supply of new houses, but prices kept its level or even higher, so the turnover decreased, which, in overall, elicited the keep soaring monthly supply.

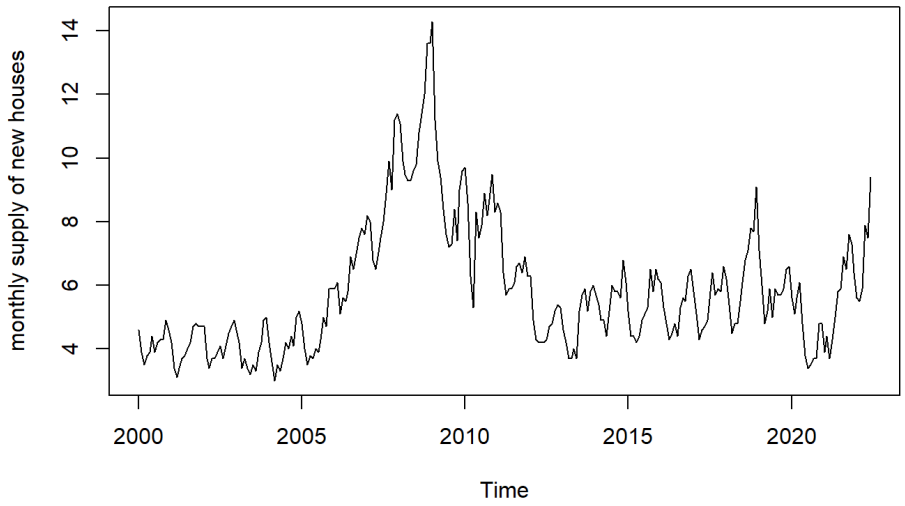

Figure 1 Monthly supply of new houses in the USA:2000-2020
Photo credit: Original

A lower number of monthly supply means that buyers are dominating the market and there are relatively few sellers; a higher number means there are more sellers than buyers. So it is obvious if the market is favoring buyers or sellers [6].

The monthly newly confirmed cases represents the severity of the pandemic in United States which keep exerting an impact on monthly supply of new houses. As it became severer, government would adopt more dour and active prevention measures, like most non-essential retail stores, including shopping malls, restaurants, barber shops and bars, were required to suspend business. Such shops that had lost their revenue stream can maintain in the short term, but if the pandemic lasted for a long time, the shopkeepers would save costs through layoffs, and in the long run, they would even face the closure[7], which caused the unstable financial income in a family and thus influence the monthly supply.

2.2.Unit Root Test

In ARMA model, in order to figure out the causal relationship between the monthly supply itself in different time points, the term should be stationary in the first place. The Augmented Dickey-Fuller Test Unit Root Test in R studio can test if the logged data series is stationary, if not, the differentiation should be applied to stabilize it. We applied two functions here, the first one is adf.test which is used to test if the series has a unit root by providing the p-value and the null hypothesis: the series has a unit root. The second one is ur.df function, it provides comparison between the test-statistics and critical values suggests whether the series is stationary or not.

The unit root test of logged series has p-value equals to 0.5196(>0.05), indicates accepting the null hypothesis, so it has unit root, the series is not stationary (please see Table 1). Similarly, its test-statistic equals to -0.0159 which is larger than all critical values, also indicates it is none-stationary. For it has a strong seasonality, the seasonal differentiation is required first, after which the unit root test has a test-statistic (-2.4998) larger than 5% and 10% critical values and only slightly smaller than 1% critical values, so we took another first differentiation to make the result more accurate. After two different differentiations, the test-statistic equals to -6.1117 which is significantly smaller than all critical values, moreover, the p-value is 0.0449(<0.05), indicates refusing the null hypothesis, so the series has already been stationary thus can be used to construct the ARMA model.

Table 1 ADF test

Variables	t-statistic	p-value
Raw series	-0.0159	0.5196
Yield series	-6.1117	0.0449**

Critical values: 1%(-2.60), 5%(-1.95), 10%(-1.62)

2.3. ARIMA Model and ARIMAX Model

We built ARIMA model based on data from January 2000 to December 2019, chose model with the least AIC value, then applied it to predict the monthly variation from January 2020 to June 2022, and this forecasting value will be mentioned as Fitted value 1 in the following passages. When monthly supply equals to zero, the logarithm may create NA value, by adding 1 to monthly supply before logarithm transition can solve this problem. The formula of ARIMA model is:

$$\log(\text{monthly housing supply} + 1) = \text{AR term} + \text{MA term} + c \tag{1}$$

Where: c is a constant. Considered the effect of pandemic, we added the monthly confirmed cases as an exogenous variable to the anterior ARIMA model to construct the ARIMAX model to fit the variation under pandemic from January 2020 to June 2022, and computed this fitted value which named Fitted value 2 in this time period. Since the monthly confirmed cases can be zero either, plus 1 before logarithm, take $\log(\text{confirmed cases} + 1)$ as the exogenous variable. The formula of ARIMAX model is:

$$\begin{aligned}\log(&\text{monthly housing supply} + 1) \\ &= \text{AR term} + \text{MA term} + \log(\text{confirmed cases} + 1) + c\end{aligned} \tag{2}$$

Where: c is a constant.

We thus created a counterfactual framework. In order to figure out if the monthly confirmed cases have an impact on the monthly housing supply, the ARIMA model provides us a result without the influence of pandemic, which is contrary to the fact. The ARIMAX model takes the monthly confirmed cases' influence into consideration, so this is the factual part. If the factual and the counterfactual part has a significant difference in statistics, then the monthly confirmed cases has a significant influence to the variation of monthly housing supply.

3. Empirical Results and Analysis

3.1. Order Selection of ARIMA Model

According to the unit root test, the model ARIMA(p,d,q)(P,D,Q) has d equals to1 and D equals to 1. In the ACF plot and PACF plot of the differentiated logged series, we used ACF to determine the order of Moving Average and PACF for Auto Regression. The PACF shown in Figure 1 is suggestive of an AR (1) model and the ACF suggests a MA (1). So an initial candidate model is an ARIMA (1,1,1). There are no other obvious candidate models. We fit an ARIMA (1,1,1) model along with variations including ARIMA (2,1,1), ARIMA (1,1,0), ARIMA (1,1,2), etc. Of these, the ARIMA (2,1,1) has a slightly smaller AIC value. In addition, it shows a seasonality in PACF, for auto correlation values are significant with lag=12 and lag=24, so there should be a seasonal auto regression, so the value of P is 1. The order of ARIMA model should be (2,1,1) or (1,1,0).

The ACF plot of the residuals from the ARIMA (2,1,1) model shows that all auto correlations are within the threshold limits, indicats that the residuals are behaving like white noise. A portmanteau test returns a large p-value, also suggesting that the residuals are white noise (Figure 2).

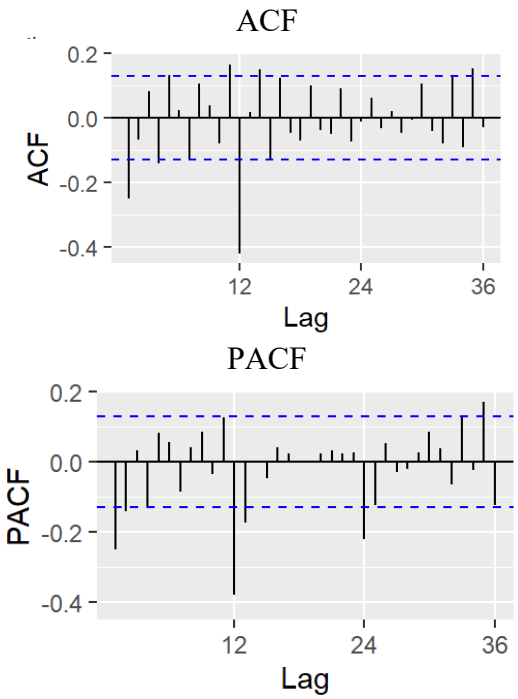

Figure 2 PACF and ACF
Photo credit: Original

3.2. Comparison between Actual Value and Fitted Value

Figures 3 and 4 depict how an actual value and two fitted values varied. Under the counterfactual framework, disparities between the blue curve and the red one in Figure 3 represent the influence of pandemic. After adding the influence of monthly confirmed cases as an exogenous variable in ARIMAX model, as it shown in Figure 4, the variation of Fitted value 2 is almost in consistent with the actual data. Intuitively, the influence from monthly confirmed cases is significant.

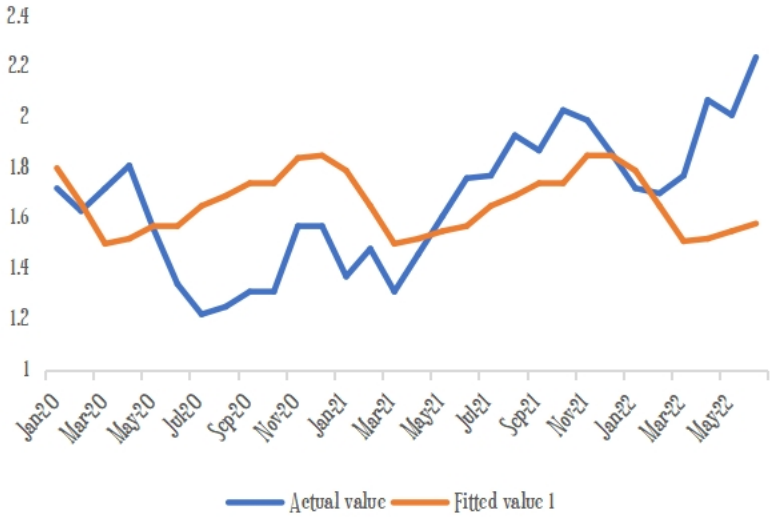

Figure 3 Actual and fitted value, model 1

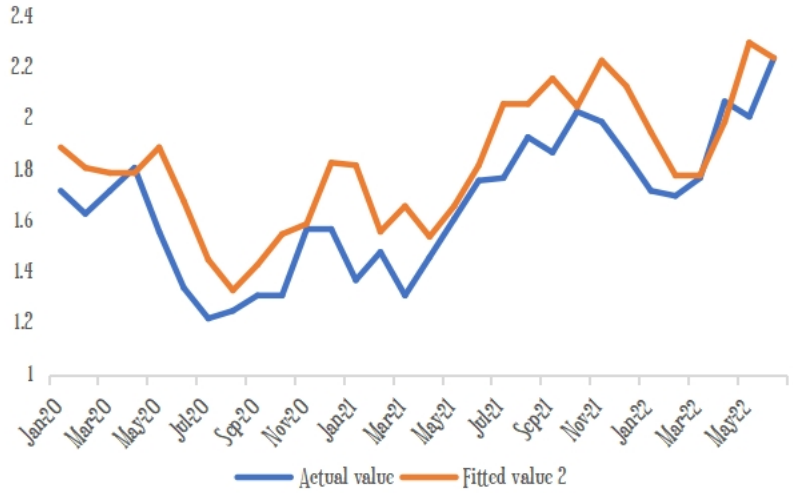

Figure 4 Actual and fitted value, model 2
Photo credit: Original

3.3. T-test

Used Double Sample T-test to examine the degree of deviation between the Actual value and Fitted values. T-test provided the p-value and the null hypothesis: true difference in means is equal to 0. Testing the two populations' compliance with the normal distribution and the consistency of their variances using F-test is required prior to each T-test. The T-test only qualifies as a student T-test when the variance between two values is constant. In Model 1, we divided the time period into two from the intersection, including March 2020 to May 2021(Period 1) and June 2021 to June 2022(Period 2), then observed the T-test results respectively. According to the result of Period 1, it has p-value equals to 0.0024(<0.05), indicates rejecting the null hypothesis, the difference in means is not equal to zero. For Period 2, the p-value is 0.0003(<0.05), so their means are quite different. In Model 2, t-test entails the time period from March 2020 to June 2022, the p-value suggests that there are differences between means. Additionally, although all three of the p-values are less than 0.05, Fitted Model 1's p-value is much smaller than that of Fitted Value 2, which equals to 0.0202, indicating that the disparity between Fitted Value 1 and Actual Value is greater. As a result, the ARIMAX model's fitting effect is better, the COVID-19 indeed had a significant impact on the monthly supply of new houses. In addition, the model 2 has the least T statistic which is -6.6722, as it shown in Table 2, suggesting there are less differences in means between the Fitted value 2 and the true value. When the parameters of the population are estimated by the statistics of the sample, the degree of freedom refers to the number of data that are independent or can changed freely in the sample. The larger the df value is, the closer the fitted value get to the actual value and the fitted value is more reliable.

Table 2 T-test

Model	Variables	T statistic	p-value	df
Model 1	Actual value Fitted value 1(Period 1)	-3.3338	0.0024***	14
	Actual value Fitted value 1(Period 2)	4.1951	0.0003***	22
Model 2	Actual value Fitted value 2	-6.6722	0.0202**	29

4. Discussion

Christopher J. Waller [8] selected the Zillow Residential Value Index, and then analyzed the US housing price and the rent used or estimated by individuals or households, pointing out the impact of changes in loan interest rates on US housing prices under the influence of COVID-19**.** Huang Xiaosi [9] used the SWOT method to explain the advantages and disadvantages of the post epidemic real estate economy, and then provided reasonable suggestions for the post epidemic real estate development strategy on the opportunities and threats it faced. The article on American real estate price index [10] shows that American homeowners are reluctant to sell their houses by analyzing the record equity level, which indirectly leads to the urgent demand for new houses in the United States. In this paper, the monthly housing supply is used to judge the time required for new housing sales, which is different from the price used to measure the quality of housing sales. This method can provide quantitative enlightenment for developers to control the number of houses and maximize their income. At the same time, we can make a relatively stable judgment on the monthly supply changes under the same environmental influence, so that developers can predict and avoid the economic losses caused by the same type of factors in advance.From the df value obtained from the ADF experiment in Table 2, it can be seen that COVID-19 has a negative impact on the monthly supply of new houses, and from Figure 1, it can be seen that the monthly supply of new houses is continuously decreasing, and the average house price in the United States is continuously increasing during the epidemic, which is an unreasonable situation.

There are some limitations and drawbacks in the data analysis process. The monthly confirmed cases can not represent the whole influence that the pandemic had on the monthly supply. For the influence come from many other aspects. Our model just entails one part which uses the size of the confirmed cases to stand for the severity of the pandemic. In addition, the ARIMA model needs the data to be stationary in the first place, before building the model, so we used differentiation to make it stationary, so essentially it can only capture the linear relation, instead of some non-linear feature, which may be the reason why the forecasting curve does not fit well in sharp turning points in time series. There are many other causes for the variation of the monthly supply and it is hard to find out and get the exact data.

5. Conclusion

Based on the new house supply data of the United States for 22 years, this paper judges the negative impact of COVID-19 on the real estate industry in the United States by establishing the difference in DF value obtained by ARMA model, discusses the possible reasons for the fluctuation of new house supply in the United States, and puts forward some suggestions on how to deal with the crisis of the real estate industry in similar situations in the future, so as to avoid the economic downturn of real estate enterprises in similar situations and achieve the stability of the housing economy. We speculate that on the one hand, due to the impact of the epidemic situation, the construction personnel could not return to their posts and could not carry out normal construction, which led to the extension of the construction period and the insufficient number of houses that could be handed over within the specified time. On the other hand, the epidemic has led to massive unemployment in the United States and the closure of physical sales centers. The existing houses cannot be digested, and the return of funds cannot be guaranteed. However, the project funds must be paid on time according to the contract. At the same time, more than 90% of the capital of real estate companies comes from bank loans. Investors will also face higher liquidated damages and payment of due principal and interest. As a result, companies that entered the website were delayed, while companies that did not enter the website were expected to close down, which directly led to the decline in housing supply. From the perspective of consumers, affected by the epidemic, the national unemployment rate in the United States continues to rise, and those who are not unemployed have not been paid in time. From the source, they cannot pay high rents or loans when buying houses. Some people even sell their existing houses at high prices to get cash, forming a vicious circle. Therefore, combined with the ARMA

model in this paper to analyze the impact of COVID-19 on the real estate industry, and combined with the housing supply in the United States, we give advice for consumers, investors and policies:

(1) For consumers, grasp the employment opportunities, and use the policy to reduce the interest rate of housing loans and the reasonable loan extension period provided by banks to alleviate their own pressure.

(2) For investors, accelerate the recovery of the project progress, offset the project delay caused by the epidemic, adjust the house price appropriately according to market changes and provide consumers with more affordable purchase plans.

(3) In terms of policy, because the main factor of inflation in the United States is the high house price, the adjustment of the mortgage interest rate will become the most important means to control the house price and sales volume in the United States. To a certain extent, the adjustment of the interest rate will have an impact on the house price and sales volume.

References

[1] Petros Sivitanides. Monthly Supply of Homes: Formula and Fluctuations in USA[EB/OL]. (2020-02-01) [2022-09-25]. https://property-investment.net

[2] FRED economic data. Monthly Supply of New Houses in the United States (MSACSR)[DB/OL]. [2020-09-25]. https://fred.stlouisfed.org/graph/?g=U7k5

[3] Category Archives: Housing Supply. (2022-09-20) [2022-09-25]. Housing Supply-Realtor.com Economic Research

[4] Petros Sivitanides. Monthly Supply of Homes: Formula and Fluctuations in USA. (2020-02-01) [2022-09-25]. https://property-investment.net

[5] Dana Anderson. Demand For Vacation Homes Drops to Lowest Level in Nearly Two Years. (2022-03-15) [2022-09-25]. Demand For Vacation Homes Drops to Lowest Level in Nearly Two Years (redfin.com)

[6] Months of Supply. [2022-09-25]. www.redfin.com

[7] Charles S. Gascon , Jacob Haas. The Impact of COVID-19 on the Residential Real Estate Market. (2020-10-06) [2022-09-25]. The Impact of COVID-19 on the Residential Real Estate Market | St. Louis Fed (stlouisfed.org)

[8] Christopher J. Waller, Xu Hengge. The Hot Real Estate Market: The Impact of the Epidemic on the US Real Estate Market from a Policy Perspective. Financial Market Research, 2022 (06): 83-88

[9] Huang Xiaosi. Analysis of Real Estate Industry in the Post epidemic Period Based on SWOT Method. Sichuan Building Materials, 2022,48 (01): 51-52+54

[10] First American Real House Price Index: Record Equity Levels Unlikely to Spur Many Homeowners to Sell. Manufacturing Close - Up, 2021,