Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/20/20230217

Covid-19's Influence on the Sale of Chinese Commercial Residential Housing

—Evaluation Analysis Based on Multi Variable Linear Regression Model

Jiaming Yang[1,a,*] , and Lei Chang[2]

[1]School of Letters and Science, University of California, Santa Barbara, Goleta, California, 93117
[2]School of Big Data, Changzhou University, Changzhou, Jiangsu, China, 213159
a. Jiaming_yang@ucsb.edu
**corresponding author*

Abstract: In January 8th, 2023, Chinese government officially terminated its quarantine policies after the publication of "The Overall Plan of Implementing Category B Management Measures to Covid-19 As a Category B Infectious Disease", since then, Chinese society entered into a brand new phase of the recovery of economic production. Next, in January 17th, 2023, the vice prime minister He Liu made a speech about his optimistic opinions about future Chinese Housing Market in the annual meeting of Davos Forum. This paper chooses the sale of Chinese commercial residential housing as the research object and acquires annual data of Chinese Housing Market since 1998 from Huibo terminal platform and Qianzhan database as the experimental data. This paper uses STATA to model experimental data and conducts research of the sale of Chinese commercial residential housing through the method of linear regression analysis. Bases on the analysis of the experimental data, this paper finds out the conclusion that covid-19 causes significant impacts toward the sale of Chinese commercial residential housing. Bases on the fitting result derived from the fitting analysis of the sale of Chinese commercial residential housing from 2022 to 2024 in the experimental model, it points out that there is going to be significant recovery in Chinese housing market from 2023 to 2024.

Keywords: China, sale area of commercial housing, Covid-19, sale of Chinese commercial residential housing, multiple linear regression model

1. Introduction

Covid-19 pandemic is the second global-scale epidemic disease in 21st century after Influenza A virus subtype H1N1 in 2009. Covid-19 causes significant negative impact to the economic development of the global society. Housing market, as a Capital-intensive industry in the Tertiary sector of the economy, plays a crucial role in several aspects such like citizens 'happiness, the level of urban construction development, and sustainable economic development, etc. Currently, most researchers in the field of housing market focus on the research of the trend of the national economic development, national policy, technic development, and market demand. However, there is not too much research progress about the sale of housing market in the past, it is of some reference value to

DOI: 10.54254/2754-1169/20/20230217

conduct scientific research of the sale of Chinese commercial residential housing for the research field in the housing market. The sale of Chinese commercial residential housing is a very import indicator that reflects the sale performance of Chinses housing market, as a result, this paper is a national medium range forecast of the national housing market within next 5 years based on the sale of Chinese commercial residential housing and the future perspective of sale. This paper takes methodology as the main research method, and it takes the continuity, relativity, and analogy as the theoretical foundation. The sale of Chinese commercial residential housing is selected as dependent variable, and national individual income tax, gross domestic product, national social practitioners, average wage of staff and workers employed, national land space purchased in this year, national sale price of commercial residential housing are selected as control variables. The experiment in this paper is a multiple linear regression analysis of the experimental data based on the experimental model derived from STATA, the Covid-19 period is selected as the time split point of the experiment. The first part of the experiment result is the result of the multiple linear regression of the experimental data ranges from 1998 to 2019, and the next part of the experiment is the comparison between the fitted value of the sale of 2020 to 2021 Chinese commercial residential housing that derived from experiment model to the real value of the sale of 2020 to 2021 Chinese commercial residential housing. The research progress of this paper will contribute to the research field of housing market research, and it will provide certain scientific evidence for the actual controllers of real estate enterprises for references when they are making operating decisions about housing market.

2. Literature Review

2.1. The Thinking and Analysis of Domestic Scholars Toward the Development of Chinese Housing Market

As an important research indicator, the development of Chinese housing market is one of the research objects for scholars when they are conducting researches of Chinese national economy and people's livelihood since the beginning of this century. Currently, most domestic scholars are conducing comparison analysis of the supply-demand side and other financial indicators in the housing market, and there is a lot of prominent research progress has been made so far. For instance, Tianyu Yang analyzes main reasons contribute to the Chinese real estate bubble through chart data of demand side and supply side [1]; An Wang evaluates the credit exposure of real estate enterprises through Z-score model [2]; Yimang Fu and his colleagues discuss the regional diversity of the Chinese real estate market among different regions through panel data model and empirical analysis method [3];Dequan Yao and his colleagues use Finkelstein's power model to analyze the significant differences of the achievements of real estate enterprises brought by the influence from owner's equity, prestige power, expert power, and structure power[4]; Biao Shi uses principal components analysis,PCA to build and analyze the financial indicators of Chinese real estate market, and he uses general panel model to examine the relationship between commercial banks' bad loan ratio and explanatory variables such like real estate financialization index, economic growth level, consumer price level, etc. [5]

2.2. The Thinking and Analysis of Foreign Scholars Toward the Development of Chinese Housing Market

As the second largest economy in the world, Chinese real estate market size is approximately about 25% to 30% of China's economic aggregate. Many overseas investment institutions joined Chinese real estate capital market in certain degrees through investing in financial products such like bond and stock. Therefore, predicting the future trend of Chinese real estate market is a very important research topic for overseas research institutions. Liang Wesley and his colleagues use fundamental research method to analyze the real estate data of Beijing and Shanghai. From the comparison of the

DOI: 10.54254/2754-1169/20/20230217

real estate data between Beijing and Shanghai and comparison between Chinese real estate supply-side data and demand-side data, Liang Wesley and his colleagues point out the trend of bubblization in Chinese real estate market [6]; Sun Weizeng and his colleagues use regression discontinuity design,RDD research method to analyze the impact of property-purchasing limitations policy to the property resale prices and property resale numbers in Beijing [7]; Edward Glaeser and his colleagues compare the condition of bubblization between United States and China and use the comparison result combined with the factors from the supply-side and demand-side of Chinese real estate market to predict China real estate market's next 20 years' real estate price trend from 2010 to 2030 [8]; Chi-Wei Su uses duration model to analyze the correlation between the Chinese commercial residential price and Chinese marriage rate. Later Chi-Wei Su indirectly proves opinions listed in the paper by researching Chinese female unemployment rate and the growth of gross domestic product [9]; Ka Shing Cheung and his colleagues use the price gradient model to analyze the transaction prices of commercial residential housing from 62 regions of Wuhan's 9 districts and Wuhan's Covid-19 data. Ka Shing Cheung and his colleagues derive their conclusion about the impact of Wuhan's commercial residential housing prices brought by Covid-19 from studies listed above [10].

3. Empirical Design

3.1. The Introduction of Experiments

3.1.1. The Introduction of the Research Methods of Experiments.

To conduct scientific and effective analysis toward the sale of Chinese commercial residential housing, this paper selects empirical analysis as the research method. Bases on the condition that the sale of Chinese commercial residential housing is the research object, the sale of Chinese commercial residential housing is the dependent variable **Y** of this paper. Because empirical analysis is the hypothesis of research object, this paper chooses 6 dataset $\mathbf{X_1}$: Chinese national individual income tax, $\mathbf{X_2}$: Chinese gross domestic product, $\mathbf{X_3}$: Chinese national social practitioners, $\mathbf{X_4}$: Chinese average wage of staff and workers employed, $\mathbf{X_5}$: Chinese national land space purchased in this year, $\mathbf{X_6}$: Chinese national sale price of commercial residential housing as the explanatory variables of the multiple linear regression model in this paper. By using multiple linear regression model, this paper derives formula of the sale of Chinese commercial residential housing (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \mu_i \quad (1)$$

which dependent variable Y is the sale of Chinese commercial residential housing, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ is the regression coefficients of the multiple linear regression model, X_1, X_2, X_3, X_4, X_5, X_6 is the 6 explanatory variables introduced above, and μ_i is the possible random disturbance term that might appear in the multiple linear regression model in this model.

Table 1: The introduction of variables.

The categories of variables	The names of variables	The signal of Variables	The significance of variables
Dependent Variables	The sale of Chinese commercial residential housing	Y	Annual cumulative data, which is the total contract price of the new built commercial residential housing of China in one year

DOI: 10.54254/2754-1169/20/20230217

Table 1: (continued).

Explanatory Variables	Chinese national individual income tax	X_1	Annual cumulative data, which is the tax that each natural person should pay to the government from a portion of their individual income
	Chinese gross domestic product,	X_2	Annual cumulative data, which is the value of products and labor from the national economy
	Chinese national social practitioners	X_3	Annual cumulative data, which is the total number of populations who are of labor capacity and might or actually attend social labor with qualified labor age.
	Chinese average wage of staff and workers employed	X_4	Annual cumulative data, which is the average wage of each employee from all the work units in the country in one year.
	Chinese national land space purchased in this year	X_5	Annual cumulative data, which is the total area of land that real estate enterprises in the country legally acquired with any methods in one year.
	Chinese national sale price of commercial residential housing	X_6	Annual cumulative data, which is the average square meter price of commercial residential housing in the country sold in one year

This paper takes multiple linear regression model to conduct analysis of experimental data, the main research objective is to examine the hypotheses and choose the optimal model to conduct fitting analysis toward the dependent variable Y from 2022 to 2024.

DOI: 10.54254/2754-1169/20/20230217

3.1.2. The Group Method of the Dependent Variable Y and Explanatory Variables.

This paper takes the analysis of time series as the statistical method. Bases on the correlation between Covid-19 and dependent variable **Y**, this paper selects the event of the breakout of Covid-19 as the time split point. The time split point is used as the method to divide dependent variable **Y** and explanatory variables into groups for the later analysis of the multiple linear regression model in this paper.

3.1.3. The Hypotheses of Experiment Results.

Hypothesis 1: The Covid-19 from 2019 to 2021 did not cause significant impact on dependent variable **Y**.

If the experiment results derived from multiple linear regression model in this paper are not significantly different from the real values of the dependent variable Y from 2020 to 2021, it points out that Covid-19 did not cause significant impact on the dependent variable Y. As a result, it will be of more effectiveness for the multiple linear regression model to acquire fitting values by adopting data from 1998 to 2021.

Hypothesis 2: The Covid-19 from 2019 to 2021 caused significant impact on dependent variable Y.

If the experiment results derived from multiple linear regression model in this paper are significantly different from the real values of the dependent variable Y from 2020 to 2021, it points out that Covid-19 caused significant impact on the dependent variable Y. As a result, it will be of less effectiveness for the multiple linear regression model to acquire fitting values by adopting data from 1998 to 2021.

3.1.4. The Concrete Method for the Demonstration of the Hypotheses.

A. The comparison of residual values.

This paper derives the residual value by subtracting the fitting values from 2000 to 2019 with observed value from dependent variable Y. Next this paper gets the average residual values from 2000 to 2019 and acquires the population standard deviation of the residual values of sample by taking the method of the calculation of the standard deviation from the sample.

B. The analysis of the probability of the sample's normal distribution

This paper takes the method of transforming normal distribution to standard normal distribution. After the average residual values and population standard deviation value are acquired, this paper separately calculates the specific value of the dependent variable Y in 2020 and 2021 after transforming from normal distribution to standard normal distribution by using the formula:

$$\frac{residual\ value - expected\ value}{standard\ deviation} \tag{2}$$

The two specific values derived from (2) are used to look for the probabilities that the residual values of 2020 and 2021 did not occur, and the probabilities are found by using normal distribution probability graph. Next, the probabilities of occurrence of the residual values of 2020 and 2021 are derived by using 1 minus the probabilities that the residual values of 2020 and 2021 did not occur. In the end, the two hypotheses are demonstrated by comparing the probabilities of occurrence of the residual values of 2020 and 2021.

DOI: 10.54254/2754-1169/20/20230217

3.2. Experimental Data

3.2.1. Data Source

This paper divides dependent variable **Y** and 6 explanatory variables from 1998 to 2024 into 27 groups of data. To keep the regression effect of the multiple linear regression model, this paper consistently chooses the annual data for the variable Y and 6 explanatory variables. The data source of this paper is Huibo terminal platform and Qianzhan database. Most of the dataset that are selected as the explanatory variables are strongly indicative factors in the aspects of society, economy, and employment. The sample of dependent variable **Y** is composed with 24 groups of annual data from 1998 to 2021, and the sample of explanatory variable is composed with 27 groups of annual data from 1998 to 2024.

3.2.2. Analysis Tool

Because the research is conducted based on multiple linear regression model, this paper selects STATA as the measurement software for the multiple linear regression model. The experimental data of the multiple linear regression model are chronologically divided into 3 time-ranges: 1998 to 2019 (Group 1), 1998 to 2020 (Group 2), 1998 to 2021 (Group 3) to better demonstrate the hypotheses of the experiments. Due to the need of the calculation of the probability of standard normal distribution, the latter part of the experiment utilizes the function of Excel to conduct the calculation.

3.3. Experimental Procedure

3.3.1. The Experimental Data Processing

This paper uses formula (1) as the research method. Before beginning running the multiple linear regression model by STATA, this paper firstly conductes the standardization of the dependent variable **Y** and explanatory variable

Table 2: The standardization of the experimental data.

Year	x5	x1	x2	x3	x4	y	x6	zee_y	zee_x1	zee_x2	zee_x3	zee_x4	zee_x5	zee_x6
1998	10109.3	338.2	84402	70637	7479	2.0e+07	1854	-.993865	-1.113802	-1.121526	-2.359993	-1.12519	-2.141209	-1.129629
1999	11958.9	412.83	89677	71394	8346	2.4e+07	1857	-.9861794	-1.070744	-1.105664	-1.982783	-1.097848	-1.94075	-.1.128534
2000	16905.2	510.18	99215	72085	9371	3.2e+07	1948	-.9708082	-1.014577	-1.076983	-1.63846	-1.065525	-1.404674	-.1.095303
2001	23409	716.01	109655	73025	10870	4.0e+07	2017	-.9554371	-.8958222	-1.04559	-1.170062	-1.018253	-.699796	-.1.070106
2002	31356.8	605.83	120333	73740	12422	5.0e+07	2092	-.9362231	-.9593913	-1.013481	-.8137801	-.9693103	.1615815	-.1.042717
2003	35696.5	567.25	135823	74432	14040	6.5e+07	2197	-.9074022	-.9816503	-.9669033	-.4689591	-.9182861	.6319153	-.1.004374
2004	39784.7	694.82	159878	75200	16024	8.6e+07	2608	-.867053	-.9080479	-.8945702	-.0862675	-.8557199	1.074992	-.8542866
2005	38253.7	837.97	184397	75825	18364	1.5e+08	2936.96	-.7440838	-.825466	-.8192181	.2251676	-.7819272	.9090629	-.7341583
2006	36573.6	981.54	216314	74978	21001	1.7e+08	3119	-.7056559	-.7426229	-.7248679	-.1968893	-.6987684	.726975	-.6676816
2007	40245.8	1273.78	265810	75321	24932	2.6e+08	3645	-.5327304	-.5740131	-.576034	-.0259737	-.5748029	1.124966	-.475599

DOI: 10.54254/2754-1169/20/20230217

Table 2: (continued).

2008	39353.4	1488.08	314045	75564	29229	2.1e+08	3576	-.6288001	-.4503714	-.430992	.0951123	-.4392955	1.028248	-.5007961
2009	31909.5	1582.54	340903	75828	32736	3.8e+08	4459	-.3021632	-.395872	-.3502303	.2266625	-.328701	.2214827	-.1783456
2010	39953	1934.3	401513	76105	37147	4.4e+08	4725	-.1868795	-.1929217	-.1679768	.3646906	-.1895985	1.093232	-.0812087
2011	40973	2421.04	471564	76420	42452	4.9e+08	4993.17	-.0908098	.087906	.0426657	.5216539	-.0223034	1.203779	.0167205
2012	35667	2327.63	519322	76704	47593	5.3e+08	5429.93	-.0139541	.0340126	.1862734	.66317	.1398199	.6287181	.176215
2013	38814.4	2612.54	588019	76977	52388	6.8e+08	5850	.274255	.1983934	.3928444	.7992049	.291032	.9698312	.3296144
2014	33383	2950.58	636463	77253	57346	6.2e+08	5933	.1589714	.3934278	.5385149	.9367347	.4473843	.3811797	.359924
2015	22811	3446.75	676708	77451	63241	7.3e+08	6473	.3703247	.6796963	.6595311	1.035397	.6332852	-.7646068	.5571192
2016	22025.3	4034.92	744127	77603	68993	9.9e+08	7203	.8698871	1.019045	.8622592	1.111138	.8146766	-.8497604	.8236977
2017	25508.3	4785.64	827122	77640	76121	1.1e+09	7614	1.08124	1.452178	1.111824	1.129575	1.039461	-.4722751	.9737852
2018	29142	5547.55	900309	77586	84744	1.3e+09	8544.11	1.465519	1.891767	1.331897	1.102667	1.31139	-.0784571	1.313439
2019	25822.3	4154.34	990865	77471	93383	1.4e+09	9287	1.657659	1.087945	1.604197	1.045363	1.583824	-.438244	1.584725
2020	25536.3	4627.27	1.0e+06	76064	97379	1.5e+09	9980	1.849798	1.360805	1.631666	-.1540358	1.709839	-.4692405	1.837792
2021	21589.9	5596.69	1.1e+06	74652	110221	1.6e+09	10396	2.094389	1.920118	1.932365	-.3593339	2.114817	-.8969488	1.989705

3.3.2. The Experimental Procedure of Three Groups of Experimental Data

This paper starts the multiple linear regression analysis of 3 groups of experimental data with chronological sequence after the standardization of the experimental data in this paper. To keep the effectiveness of the results, the 3 groups of experimental data took the following experimental procedure: Step 1, before the multiple linear regression gets started, 4 feedback data including average value, variance, maximum value and minimal value are acquired through descriptive analysis of the 3 groups of experimental data in STATA; Step 2, after the feedback results of the descriptive analysis are acquired, this paper conducts multicollinearity test to each group, and evaluates the existence of correlation within 6 explanatory variables based on the VIF value acquired from multicollinearity test; Step 3, after the complement of two steps lists above, this paper conducts ordinary least squares regressions, OLS ,with all 6 explanatory variables, and the nonsignificant explanatory variables are removed based on the evaluation of the P-values; Step 4, another ordinary least squares regression is conducted with remaining significant explanatory variables, and the formula of multiple linear regression is postulated based on the relationship between the dependent variable **Y** and explanatory variable; Step 5, the regression effects of the three groups of experimental data are evaluated through their R-square results.

3.3.3. The Demonstration of the Effectiveness of the Regression

After the experiments of 3 groups of experimental data, this paper conducts fitting analysis based on the formula derived from the multiple linear regression results through the experiments of 3 groups of experimental data, and the results of fitting values and observed values are put in the same coordinate axis to get the visual feedback of the regression effect of the 3 groups of experimental data.

DOI: 10.54254/2754-1169/20/20230217

3.3.4. Fitting Analysis

After demonstrating the hypotheses of the experiments by using the concrete method for the demonstration of the hypotheses in 3.1, this paper selects the time-range which has the optimal regression results between dependent variable **Y** and explanatory variables, and the fitting results are acquired by taking the observed values of explanatory variables from 2022 to 2024 into the optimal multiple linear regression model.

3.4. Experimental Results

3.4.1. The Experimental Results of Linear Regression

Bases on the experimental procedure of three groups of experimental data arranged in 3.3, each of 3 groups of experimental data acquires corresponding regression results.

Table 3: The results of group 1 by conducting step 1.

Variable	Obs	Mean	Std. Dev.	Min	Max
year	22	2008.5	6.493587	1998	2019
x5	22	30438.89	9420.441	10109.3	40973
x1	22	2010.196	1558.169	338.2	5547.55
x2	22	403500.2	290519.1	84402	990865
x3	22	75419.95	2092.69	70637	77640
x4	22	37646.45	26745.74	7479	93383
y	22	4.42e+08	4.28e+08	2.01e+07	1.39e+09
x6	22	4470.962	2314.197	1854	9287

Table 4: The results of group 2 by conducting step 1.

Variable	Obs	Mean	Std. Dev.	Min	Max
year	23	2009	6.78233	1998	2020
x5	23	30225.74	9260.447	10109.3	40973
x1	23	2123.982	1617.194	338.2	5547.55
x2	23	430130	311248	84402	1015986
x3	23	75404.48	2045.922	70637	77640
x4	23	40243.52	2894735	7479	97379
y	23	4.90e+08	4.78e+08	2.01e+07	1.55e+09
x6	23	4710.486	2536.063	1854	9980

Table 5: The results of group 3 by conducting step 1.

Variable	Obs	Mean	Std. Dev.	Min	Max
year	0				
x5	24	29865.91	9226.85	10109.3	40973
x1	24	2268.678	1733.233	338.2	5596.69
x2	24	457375.2	332558.7	84402	1100000
x3	24	75373.13	2006.838	70637	77640
x4	24	43159.25	31710.43	7479	110221
y	24	5.38e+08	5.22e+08	2.01e+07	1.63e+09
x6	24	4947.382	2738.405	1854	10396

DOI: 10.54254/2754-1169/20/20230217

Table 6: The results of group 1 by conducting step 2.

Variable	VIF	1/VIF
x4	952.60	0.001058
x2	717.52	0.001394
x6	192.16	0.005204
x1	21.98	0.045499
x3	14.82	0.067480
x5	4.71	0.212520
VIF	317.30	

The results of group 1 by conducting step 2 points out that the VIF values of explanatory variables X_1, X_2, X_3, X_4, X_6 are larger than 10, which means that multicollinearity exist within these 5 variables.

Table 7: The results of group 2 by conducting step 2.

Variable	VIF	1/VIF
zee_x3	2.21	0.452062
zee_x6	2.21	0.452062
Mean VIF	2.21	

The results of group 2 by conducting step 2 points out that the VIF values of explanatory variables X_3, X_6 are smaller than 10, which means that multicollinearity does not exist within these 2 variables.

Table 8: The results of group 3 by conducting step 2.

Variable	VIF	1/VIF
zee_x1	18.28	0.054713
zee_x6	18.28	0.054713
Mean VIF	18.28	

The results of group 3 by conducting step 2 points out that the VIF values of explanatory variables X_1, X_6 are larger than 10, which means that multicollinearity exist within these 2 variables.

Source	SS	df	MS	Number of obs	=	22
				F(6, 15)	=	396.51
Model	20.8684222	6	3.47807036	Prob > F	=	0.0000
Residual	.131577017	15	.008771801	R-squared	=	0.9937
				Adj R-squared	=	0.9912
Total	20.9999992	21	.999999961	Root MSE	=	.09366

zee_y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
zee_x1	.1690945	.0958147	1.76	0.098	-.0351297	.3733186
zee_x2	-1.250576	.5474577	-2.28	0.037	-2.417454	-.0836975
zee_x3	-.1986304	.0786772	-2.52	0.023	-.3663268	-.0309341
zee_x4	1.429586	.630797	2.27	0.039	.0850743	2.774098
zee_x5	.025606	.0443338	0.58	0.572	-.0688893	.1201012
zee_x6	.8142611	.2833152	2.87	0.012	.210389	1.418133
_cons	1.13e-07	.0199679	0.00	1.000	-.0425605	.0425607

Figure 1: The results of group 1 by conducting step 3.

DOI: 10.54254/2754-1169/20/20230217

The results of group 1 by conducting step 3 point out that the P-value of X_1, X_5 is larger than 0.05 among all the 6 explanatory variables, which means that X_1, X_5 in group 1 is not significant, and the reason of this result is probably because there is lack of correlation between dependent variable Y and X_1, X_5.

Source	SS	df	MS	Number of obs	=	23
				F(6, 16)	=	505.23
Model	21.8844893	6	3.64741488	Prob > F	=	0.0000
Residual	.115509318	16	.007219332	R-squared	=	0.9947
				Adj R-squared	=	0.9928
Total	21.9999986	22	.999999937	Root MSE	=	.08497

zee_y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
zee_x1	.1684252	.0897123	1.88	0.079	-.0217565	.3586069
zee_x2	-.9821814	.4969969	-1.98	0.066	-2.035768	.071405
zee_x3	-.1192803	.0511063	-2.33	0.033	-.2276208	-.0109397
zee_x4	1.158868	.5868057	1.97	0.066	-.085104	2.402841
zee_x5	-.0006335	.0340546	-0.02	0.985	-.0728261	.0715591
zee_x6	.7401386	.2767943	2.67	0.017	.1533608	1.326916
_cons	2.72e-07	.0177168	0.00	1.000	-.0375576	.0375582

Figure 2: The results of group 2 by conducting step 3.

The results of group 2 by conducting step 3 point out that the P-value of X_3, X_6 is less than 0.05 among all the 6 explanatory variables, which means that X_3, X_6 in group 2 is significant, and the remaining explanatory variables X_1, X_2, X_4, X_5 are non-significant. The reason of this result is probably because there is lack of correlation between dependent variable Y and X_1, X_2, X_4, X_5.

Source	SS	df	MS	Number of obs	=	24
				F(6, 17)	=	526.10
Model	22.8767949	6	3.81279915	Prob > F	=	0.0000
Residual	.123204407	17	.007247318	R-squared	=	0.9946
				Adj R-squared	=	0.9928
Total	22.9999993	23	.999999969	Root MSE	=	.08513

zee_y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
zee_x1	.1506238	.0961172	1.57	0.136	-.0521658	.3534134
zee_x2	-.3857822	.4331663	-0.89	0.386	-1.299683	.5281187
zee_x3	-.0721627	.0501061	-1.44	0.168	-.1778774	.033552
zee_x4	.3031128	.4608993	0.66	0.520	-.6692997	1.275525
zee_x5	-.0225775	.0328752	-0.69	0.501	-.0919381	.0467831
zee_x6	.9732462	.2644118	3.68	0.002	.415386	1.531106
_cons	2.20e-08	.0173773	0.00	1.000	-.0366629	.036663

Figure 3: The results of group 3 by conducting step 3.

DOI: 10.54254/2754-1169/20/20230217

The results of group 3 by conducting step 3 point out that the P-value of X_6 is less than 0.05 among all the 6 explanatory variables, which means that X_6 in group 3 is significant, and the remaining explanatory variables X_1, X_2, X_3, X_4, X_5 are non-significant. The reason of this result is probably because there is lack of correlation between dependent variable Y and X_1, X_2, X_3, X_4, X_5. In account of the condition that this paper takes multiple linear regression model, and X_1 is the explanatory variable in group 3 that has P-value only larger than X_6. As a result, X_1 is considered as a significant explanatory variable in group 3.

Source	SS	df	MS	Number of obs	=	24
				F(4, 19)	=	13.73
Model	17.0869674	4	4.27174185	Prob > F	=	0.0000
Residual	5.91303378	19	.311212304	R-squared	=	0.7429
				Adj R-squared	=	0.6888
Total	23.0000012	23	1.00000005	Root MSE	=	.55786

zee_y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
zee_x2	4.551507	3.208899	1.42	0.172	-2.164795	11.26781
zee_x3	.1274222	.1763057	0.72	0.479	-.2415898	.4964343
zee_x4	-7.444678	3.65235	-2.04	0.056	-15.08913	.1997785
zee_x6	3.597944	1.632587	2.20	0.040	.1809011	7.014988
_cons	4.77e-07	.1138735	0.00	1.000	-.2383395	.2383405

Figure 4: The results of group 1 by conducting step 4 and step 5.

Bases on the feedback result from figure 4, it points out that the R-square score between dependent variable Y and 4 significant variables from 1998 to 2019 is 74.29%.

Source	SS	df	MS	Number of obs	=	23
				F(2, 20)	=	1196.99
Model	21.817727	2	10.9088635	Prob > F	=	0.0000
Residual	.182271581	20	.009113579	R-squared	=	0.9917
				Adj R-squared	=	0.9909
Total	21.9999986	22	.999999937	Root MSE	=	.09547

zee_y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
zee_x6	1.085796	.0302715	35.87	0.000	1.022651	1.148941
zee_x3	-.1264226	.0302715	-4.18	0.000	-.1895679	-.0632773
_cons	1.89e-07	.0199058	0.00	1.000	-.0415227	.041523

Figure 5: The results of group 2 by conducting step 4 and step 5.

Bases on the feedback result from figure 5, it points out that the R-square score between dependent variable Y and 2 significant variables from 1998 to 2019 is 99.17%.

DOI: 10.54254/2754-1169/20/20230217

Source	SS	df	MS	Number of obs	=	24
				F(2, 21)	=	815.52
Model	22.7076318	2	11.3538159	Prob > F	=	0.0000
Residual	.292367471	21	.013922261	R-squared	=	0.9873
				Adj R-squared	=	0.9861
Total	22.9999993	23	.999999969	Root MSE	=	.11799

zee_y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
zee_x1	.0880991	.1051832	0.84	0.412	-.1306413	.3068395
zee_x6	.907755	.1051832	8.63	0.000	.6890146	1.126495
_cons	1.08e-08	.0240851	0.00	1.000	-.0500878	.0500878

Figure 6: The results of group 3 by conducting step 4 and step 5.

Bases on the feedback result from figure 6, it points out that the R-square score between dependent variable Y and 2 significant variables from 1998 to 2019 is 98.73%.

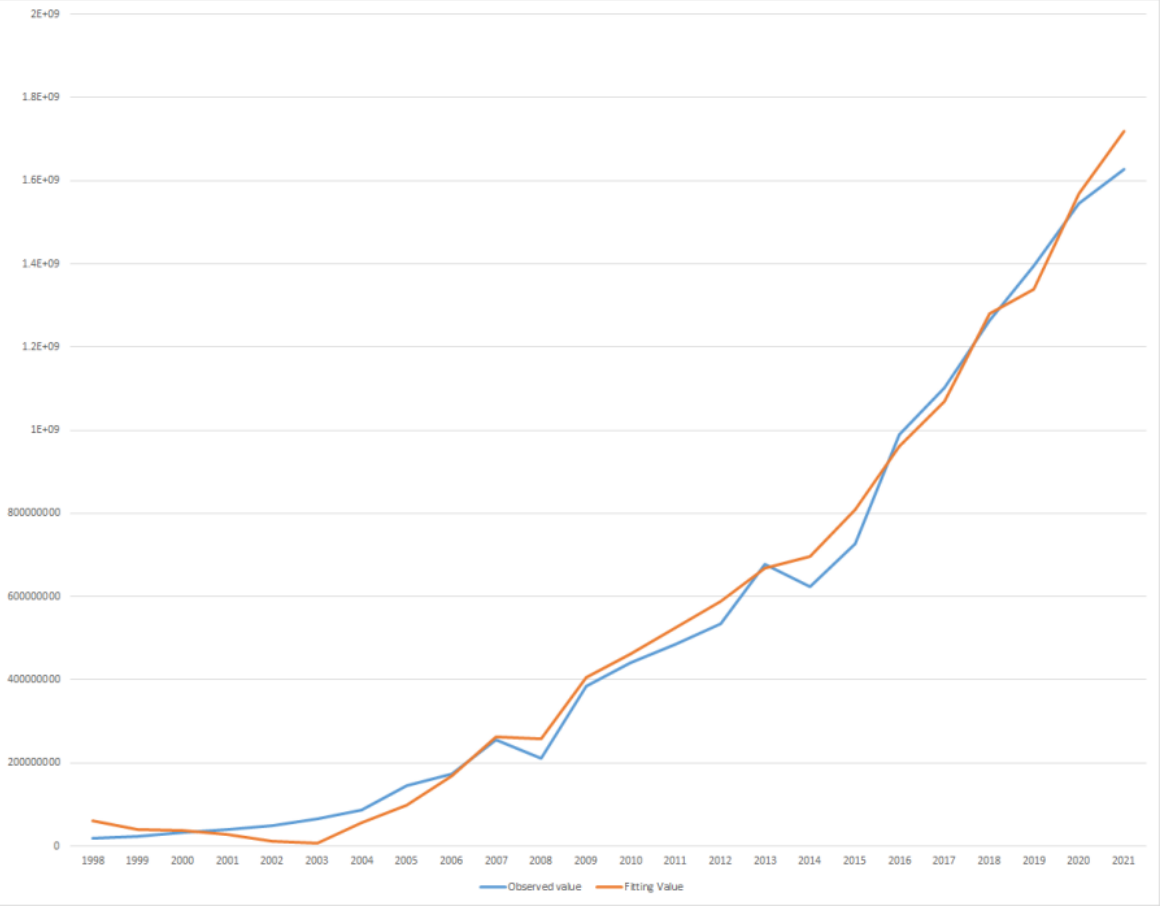

Figure 7: The comparison between observed values and fitting values.

DOI: 10.54254/2754-1169/20/20230217

From the comparison result between dependent variable Y's fitting value and observed value, it points out that the regression effect of the multiple linear regression model is relatively fair, which it reflects the observed condition of dependent variable Y in the real world correctly.

3.4.2. The Demonstration of Hypotheses

In the concrete method for the demonstration of the hypotheses in 3.1, this paper introduces the method for the demonstration of hypotheses, and according to the experimental results of linear regression, table 9 lists the observed value, fitting value, residual value of dependent variable **Y** from 1998 to 2021 and get the following results below:

Table 9: Observed value, Fitting value, Residual value (Unit: 10,000 Yuan).

Year	Observed value	Fitting value	Residual value
1998	20068676	61700000	41631324
1999	24137347	40800000	16662653
2000	32286046	38000000	5713954
2001	40211543	29100000	-11111543
2002	49578501	13000000	-3657801
2003	65434492	6271745	-59162747
2004	86193667	56400000	-29793667
2005	145637616	99100000	-46537616
2006	172878100	169000000	-3878100
2007	255658100	263000000	7341900
2008	211960000	259000000	47040000
2009	384329000	405000000	20671000
2010	441206500	462000000	20793500
2011	486194000	526000000	39806000
2012	534670000	589000000	54330000
2013	676949367	669000000	-7949367
2014	623960000	696000000	72040000
2015	727530000	810000000	82470000
2016	990641734	962000000	-28641734
2017	1102395075	1070000000	-32395075
2018	1263930000	1280000000	16070000
2019	1394399683	1340000000	-54399683
2020	1545669552	1570000000	24330448
2021	1627299000	1720000000	92701000

Bases on the value provides in table 9, this paper calculates the average residual value α from 2000 to 2019, sample average variance β. Next, take 24330448, the residual value of 2020 into formula (2) to get the distribution value ω that the value of normal distribution less than the residual value of 2022, which is written as:

$$If\ X \sim N(\mu, \sigma^2), then, Z = \frac{X-\mu}{\sigma} \sim N(0,1) \quad (3)$$

which X is expected value, μ is average value, and σ is standard deviation.

DOI: 10.54254/2754-1169/20/20230217

As a result,

$$\omega = \frac{24330448-\alpha}{\sqrt{\beta}} \approx 0.53 \quad (4)$$

Similarly, this paper takes 92701000, the residual value of 2021 in formula (2) to get the distribution value b that the value of normal distribution less than the residual value of 2022, which is written as:

$$b = \frac{92701000-\alpha}{\sqrt{\beta}} \approx 2.22 \quad (5)$$

The hypotheses of this paper is to discuss whether Covid-19 caused significant impact on dependent variable Y, and Covid-19 is a time-based continuous event. As a result, it is necessary to use ω to find out P(a), the probability that the residual value in the normal distribution is larger than the actual residual value in 2020 and use b to find out P(b), the probability that the residual value in the normal distribution is larger than the actual residual value in 2021, which is written as:

$$P(c) = 1 - P(a) = 1 - 0.7019 = 0.2981 = 29.81\% \quad (6)$$

$$\mathrm{P(d)} = 1 - \mathrm{P(b)} = 1 - 0.9868 = 0.0132 = 1.32\% \quad (7)$$

From the comparison of the value P(c) and P(d), it comes out that P(c) is 29.81%, which is the event with a fair probability of occurrence, and P(d) is 1.32%, which is a small probability event. As a result, hypothesis 2 in this paper is true, that Covid-19 from 2019 to 2021 caused significant impact on dependent variable Y. The reason that P(c) is the event with a fair probability of occurrence is probably because the impact of Covid-19 to dependent variable Y is continuous and progressive, an such impact finally caused significant and negative impact on dependent variable Y and get proved by finding out that P(d) is a small probability event.

3.4.3. Fitting Value from 2022 to 2024

This paper uses fitting analysis designed in 3.3, the multiple linear regression models derived from the experimental results of linear regression and the true hypothesis proved in 3.4 as the foundation. After consideration, this paper chooses multiple linear regression models derived from the experimental data of group 1 to conduct fitting analysis and get the results below:

Table 10: Observed value from 1998 to 2021, Fitting value from 1998 to 2024 (Unit: 10,000 Yuan).

Year	Observed Value	Fitting value
1998	20068676	61700000
1999	24137347	40800000
2000	32286046	38000000
2001	40211543	29100000
2002	49578501	13000000
2003	65434492	6271745
2004	86193667	56400000
2005	145637616	99100000
2006	172878100	169000000
2007	255658100	263000000
2008	211960000	259000000
2009	384329000	405000000

DOI: 10.54254/2754-1169/20/20230217

Table 10: (continued).

2010	441206500	462000000
2011	486194000	526000000
2012	534670000	589000000
2013	676949367	669000000
2014	623960000	696000000
2015	727530000	810000000
2016	990641734	962000000
2017	1102395075	1070000000
2018	1263930000	1280000000
2019	1394399683	1340000000
2020	1545669552	1570000000
2021	1627299000	1720000000
2022	E	1870000000
2023	E	2010000000
2024	E	2140000000

Annotation: E means this data has not been published by National Bureau of Statistics of China yet.

This paper comes out the conclusion from the results in Table 4 that since China has entered post Covid-19 era, dependent variable Y will gradually returned to the status of 2019, the time before 2019. According to the fitting values from 2022 to 2024, it points out that dependent variable Y will perform a much faster year-on-year ratio raising speed compared with the year-on-year ratio raising speed during Covid-19 period. This conclusion will play a positive reference role in the Chinese national economy and people's livelihood and real estate enterprises.

4. Conclusion

This paper takes the method of empirical analysis, and uses STATA to build multiple linear regression model and analyze whether the Covid-19 from 2019 to 2022 caused significant impact on the sale of Chinese commercial residential housing. This paper also acquires the fitting value of the sale of Chinese commercial residential housing from 2022 to 2024. Through the empirical analysis, this paper comes out the conclusion that Covid-19 caused significant negative impact on the sale of Chinese commercial residential housing. Bases on the fitting analysis of the sale of Chinese commercial residential housing from 2022 to 2024 by using multiple linear regression models, this paper comes out the conclusion that the sale of Chinese commercial residential housing will perform a significant faster year-on-year ratio raising speed compared with the speed during the period of Covid-19. There is a lot of improvement space of this paper. From the experimental results of this paper, it points out that the multicollinearity of explanatory variables is relatively high and the correlation between dependent variable Y and explanatory variables is relatively low, which a lot of data with potential research value did not indicate on the experimental data of this paper. In the future the research of Chinese real estate market can focus more on the aspect of market sales. Moreover, there are many other important indicators related with market sale besides the sale of Chinese commercial residential housing for researches to select and conduct pertinent analysis.

DOI: 10.54254/2754-1169/20/20230217

References

[1] *Tianyu Yang. The research of the bubblization issue in Chinese real estate market. [j]. The working paper of Institute of Economic Research of Renmin University of China 1-19.*

[2] *An Wang. The analysis of real estate's credit exposure based on Z-score model. [J]. Finance and Management, 2020(03): 1-4.*

[3] *Yimang Fu, Xiaohu Li, Chenyi Zhang, Liuying Xie. Financial development scale, threshold effect and real estate inventory—empirical analysis based on threshold panel model. [J]. Financial Theory & Practice, 2022(01): 35-42.*

[4] *Dequan Yao, Danyu Wen. The research of the influence of management power to real estate enterprise performances—the regulative effect based on the ownership concentration. [J] Financial Theory & Practice, 2022(05):100-106.*

[5] *Biao Shi. The uncertainty of economic policies, the financialization of real estate market and the commercial banks' bad loan ratio—the empirical research based on panel data of provinces. [J] Exploration of Financial Theory, 2022(10):1-16.*

[6] *Liang Wesley, and Gary Smith. "China Housing Market: Boom or Bubble?" Pomona College, 2019, http://economics-files.pomona.edu/GarySmith/.*

[7] *Sun, Weizeng et al. "The Housing Market Effects of Local Home Purchase Restrictions: Evidence from Beijing." The Journal of Real Estate Finance and Economics 55, 3 (October 2017): 288-312 © 2016 Springer Science+Business Media New York*

[8] *Glaeser, Edward, Wei Huang, Yueran Ma, and Andrei Shleifer. 2017. "A Real Estate Boom with Chinese Characteristics." Journal of Economic Perspectives, 31 (1): 93-116.*

[9] *Chi-Wei Su, Khalid Khan, Lin-Na Hao, Ran Tao & Adelina Dumitrescu Peculea (2020) Do house prices squeeze marriages in China?, Economic Research-Ekonomska Istraživanja, 33:1, 1419-1440, DOI: 10.1080/1331677X.2020.1746190*

[10] *Cheung, Ka Shing, et al. "Housing Market in the Time of Pandemic: A Price Gradient Analysis from the COVID-19 Epicentre in China." Journal of Risk and Financial Management, vol. 14, no. 3, Mar. 2021, p. 108. Crossref, https://doi.org/10.3390/jrfm14030108.*