ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

Research on the Influencing Factors of China's Commercial Housing Sales

-- Econometric Analysis based on Provincial Panel Data

Xingzhou Wang

School of science, Minzu University of China, Beijing 100081, China

Abstract

In recent years, China's real estate industry has experienced rapid development. The real estate industry is closely related to the livelihood of the people. At the same time, the rapid growth of the real estate industry contributes to the development of related industries. In order to conduct an in-depth analysis of the development of China's real estate industry and the factors influencing the sales of commercial housing, this study utilizes a panel dataset composed of 31 provinces, regions, and cities in mainland China from 2010 to 2022 for research and analysis, aiming to explore the influencing factors of the sales of commercial housing in China. Through model selection and parameter estimation, as well as diagnostic tests like LM test and Hausman test, model comparisons were conducted. Ultimately, the fixed effects model with time effects (FE_TW) was selected as the model for analyzing the panel dataset of the real estate industry. It was concluded that regional Gross Domestic Product, the real estate development investment, the sales area of commercial housing and the total output value of construction industry significantly influence the sales of commercial housing.

Keywords

Real Estate Industry; Sales of Commercial Housing; Panel Data; Influencing Factors; Econometric Analysis.

1. Introduction

In recent years, China's economy has experienced rapid growth, and the real estate industry plays a pivotal role in the economy. The swift expansion of China's economy has provided strong impetus for the development of the real estate industry. Stable economic growth and the urbanization process have led to population mobility and urban expansion, thereby increasing the demand for housing and commercial real estate. Furthermore, the ongoing urbanization process in China, with rural populations migrating to cities, has further amplified the demand for housing and commercial real estate. The extensive development of urbanization projects and the construction of new cities have propelled the flourishing growth of the real estate industry. The sales of commercial housing refer to the total amount of real estate sales, including various types of residential and commercially used properties, in the market. This indicator is commonly used to measure the level of activity in the real estate market and the contribution of the real estate industry to the economy. Based on the fluctuations and trends in the sales of commercial housing, the government can formulate corresponding policies and measures to maintain the stable operation of the real estate market and promote sustainable economic development.

This study employs panel data for research and analysis. Panel data analysis, widely used in econometrics, is primarily utilized to analyze datasets with both time and individual (or regional) dimensions. It offers several advantages over time series or cross-sectional analysis.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

Firstly, panel data takes advantage of the temporal dimension to observe the changing trends of variables over time, enabling the analysis and control of time-related factors on the research variables. Additionally, panel data effectively controls the heterogeneity of individuals or regions, allowing for a more accurate estimation of the relationships between variables. Furthermore, panel data analysis enhances estimation efficiency and reduces the variance of parameter estimates. By combining information from both the time and individual dimensions, a more comprehensive analysis of the data is conducted, resulting in more precise and robust estimation results.

Many scholars have utilized panel data analysis to examine real-world issues, yielding accurate and precise conclusions, and proposing rational policy recommendations. For instance, Wu Qiutong [1] established a panel model to analyze the relationship between environmental pollution levels and economic development in Nanjing over the past 30 years. The study found a significant correlation between per capita GDP in Nanjing and the emissions of industrial gases and solid waste. Xia Jiechang and Liu Ruiyi [2] employed a panel autoregressive model to analyze the interactive relationship between digital economy, green development, and resource allocation in the tourism industry. The findings suggest that the digital economy effectively improves capital allocation in tourism, and labor force allocation will optimize green development in the short term. Fan Jinhong [3] utilized panel data from 11 provinces and cities in the Yangtze River Economic Belt for the years 2007-2020 to analyze the relationship between environmental regulation, technological innovation, and regional economic growth. The study concluded that environmental regulation, technological innovation, and economic growth all have positive spillover effects. Hua Shizhen [4] conducted an analysis based on panel data from nine provinces and regions in the Yellow River Basin for the years 2009-2020, examining the relationship between green finance and the ecological economy of the Yellow River Basin. The results demonstrate that green finance significantly contributes to the development of the ecological economy in the Yellow River Basin.

Furthermore, there exists a wealth of relevant research literature on the real estate industry, with many scholars employing various methods to analyze the influencing factors of housing prices in depth. For example, Shen Guoyun [5] analyzed the main influencing factors of residential prices in Xining from three aspects: supply, demand, and macroeconomics. The study concluded that land prices, construction costs, and investment in real estate have the most significant impact on housing prices. Hu Jianhua [6] utilized panel models to analyze China's real estate prices and related factors from 2008 to 2016, finding that factors such as the scale of real estate development loans and per capita disposable income have a significant impact. Wu Jiamu [7] used average commodity housing prices in Shanghai from 2003 to 2017 and established a linear regression model, ultimately concluding that factors such as residents' savings significantly influence housing prices. Xu Lu [8] established a multiple regression model to explore the factors affecting changes in the national average housing price. The results indicate that factors such as per capita disposable income and the sales area of national commodity housing have a significant impact on housing prices.

The following structure of the paper is arranged as follows: Section Two introduces the selected variables, dataset, and data sources; Section Three specifically delves into panel data analysis. By comparing mixed regression, fixed effects models, and random effects models, an analysis is conducted, and the Hausman test is employed to select the most suitable model; Section Four presents the research conclusions of this paper along with corresponding policy recommendations.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

2. Data and Variable Description

All data used in this article were sourced from the official website of the National Bureau of Statistics of China. The panel dataset used in the analysis covers a time span from 2010 to 2022, spanning 13 years, and includes data from 31 provinces and cities in China. The dataset comprises five explanatory variables, namely regional Gross Domestic Product (GDP), real estate development investment (invest), sales area of commercial housing (area), general budget expenditure of local finance (Expenditure), and total output value of construction industry (construction). The dependent variable is commercial property sales (sales). All variables are continuous. In addition, Year is set as the time variable, representing the time span from 2010 to 2022, while Province is used as the panel variable, indicating the specific serial numbers of each of the 31 provinces and cities. The dataset contains a total of 403 observations. To begin with, a variable definition table is constructed to gain an overview of all variables in the dataset, as shown in Table 1.

Table 1. Variable definition table

Variable Type	Variable Name	Data Description	Unit
Dependent	sales	Sales of commercial housing	Hundred Million Yuan
Explanatory	GDP	Regional Gross Domestic Product	Hundred Million Yuan
	invest	Real estate development investment	Hundred Million Yuan
	area	Sales area of commercial housing	Ten Thousand Square Meters
	expenditure	General budget expenditure of local finance	Hundred Million Yuan
	construction	Total output value of construction industry	Hundred Million Yuan
Other	Year	Time variable	/
	Province	Panel variable	/

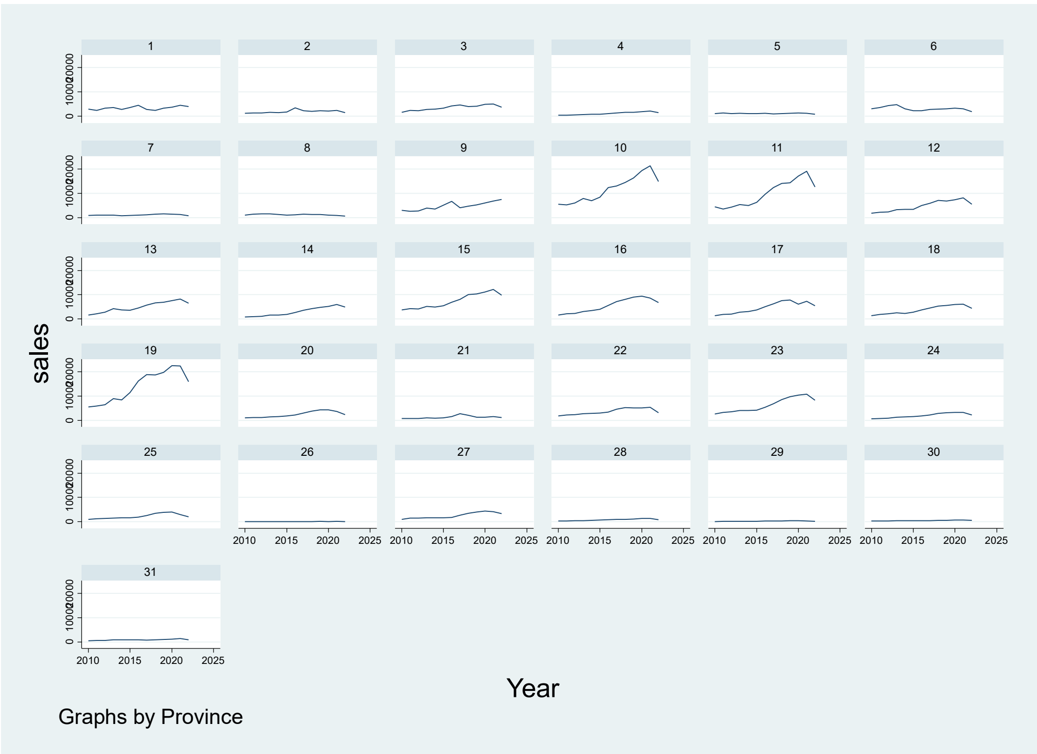

Figure 1. Time trend chart of sales in different provinces

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

Next, a time trend plot for the dependent variable "sales" across the 31 provinces and regions in China will be generated, as shown in Figure 1. From Figure 1, it can be observed that the time trends of sales for different provinces and regions in China are not entirely consistent. Some provinces and regions remain relatively stable, on the other hand, some provinces and regions show an upward trend. Therefore, to some extent, the inter-provincial differences in sales can assist in estimating the influencing factors of sales.

For panel data with relatively long time series, it is necessary to conduct a unit root test before regression to avoid potential issues of spurious regression. If the unit root test indicates the presence of a unit root, it implies that the time series is non-stationary. Modeling with non-stationary time series may lead to incorrect statistical results. Although the panel data in this study has n=31 and T=13, making it a short panel model, a panel unit root test will still be conducted for precaution. The following Fisher test will be performed. The Fisher-type unit root test is a commonly used method for unit root testing in panel data. It combines unit root test statistics for individual sequences to obtain a unit root test statistic for the panel sequence. The null hypothesis of this test is that there exists a unit root in the sequences of the panel data, indicating non-stationarity. The results of the Fisher test are shown in Figure 2.

```
Fisher-type unit-root test for sales
Based on augmented Dickey-Fuller tests

H0: All panels contain unit roots           Number of panels  =     31
Ha: At least one panel is stationary        Number of periods =     13

AR parameter: Panel-specific                Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Not included
Drift term:   Included                      ADF regressions: 3 lags

                                  Statistic      p-value

 Inverse chi-squared(62)   P       137.4244       0.0000
 Inverse normal            Z        -6.3750       0.0000
 Inverse logit t(159)      L*       -6.1945       0.0000
 Modified inv. chi-squared Pm        6.7733       0.0000

 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
```

Figure 2. Fisher test result

According to the test results in Figure 2, it can be observed that the p-values for the inverse chi-square, inverse normal, inverse logit, and modified inverse chi-square statistics are all very small, indicating significance at the 0.01 level. Therefore, we reject the null hypothesis and conclude that the sequences in the panel data are stationary. To further delve into the panel dataset, we will proceed with the modeling analysis.

3. Panel Data Modeling Analysis

3.1. Mixed Regression

As a reference for panel data modeling, a mixed regression should be conducted first. The results of the mixed regression using ordinary standard errors are presented in Figure 3. From the results, it can be observed that the coefficient of determination, R^2, is quite high at 0.9351, indicating a good fit. Regarding the parameter estimates, it can be deduced that, at the 5% significance level, except for the coefficient of expenditure and construction which has a large p-value and is not statistically significant, all other variables significantly impact the sales of residential properties.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

Source	SS	df	MS			
				Number of obs	=	403
				F(5, 397)	=	1143.72
Model	5.7996e+09	5	1.1599e+09	Prob > F	=	0.0000
Residual	402624046	397	1014166.36	R-squared	=	0.9351
				Adj R-squared	=	0.9343
Total	6.2022e+09	402	15428417.8	Root MSE	=	1007.1

sales	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
GDP	.0318442	.009913	3.21	0.001	.0123558	.0513327
invest	.9338065	.0599771	15.57	0.000	.8158941	1.051719
area	.0607479	.0263649	2.30	0.022	.0089156	.1125801
expenditure	.0319699	.0483586	0.66	0.509	-.0631011	.1270408
construction	.0051566	.0165183	0.31	0.755	-.0273177	.0376309
_cons	-753.9528	116.2797	-6.48	0.000	-982.5539	-525.3518

Figure 3. Mixed regression results (standard error)

Next, a mixed regression is conducted again using robust standard errors. The results are shown in Figure 4. By comparison, it is found that ordinary standard errors are only about half the size of robust standard errors. Additionally, due to the likely presence of serial correlation in the disturbances within a province or municipality across different periods, the assumption of independent and identically distributed disturbances underlying ordinary standard errors may lead to inaccurate estimates. Therefore, employing robust standard errors for estimating the mixed regression model is more suitable. The goodness-of-fit, as indicated by the R^2 value of 0.9351, remains high, suggesting a satisfactory fit.

Linear regression

Number of obs	=	403
F(5, 30)	=	309.59
Prob > F	=	0.0000
R-squared	=	0.9351
Root MSE	=	1007.1

(Std. err. adjusted for 31 clusters in **province**)

sales	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
GDP	.0318442	.0206199	1.54	0.133	-.0102672	.0739557
invest	.9338065	.1556687	6.00	0.000	.6158886	1.251724
area	.0607479	.0574872	1.06	0.299	-.0566566	.1781524
expenditure	.0319699	.0809498	0.39	0.696	-.1333516	.1972914
construction	.0051566	.0415195	0.12	0.902	-.0796376	.0899508
_cons	-753.9528	201.3017	-3.75	0.001	-1165.066	-342.8399

Figure 4. Mixed regression result (robust standard error)

3.2. Fixed Effects Model

Next, specific estimation results for the Fixed Effects Model are considered to address the omitted variables that do not vary with time. Initial estimates are made using within-group estimators, and robust standard errors are employed in the estimation process. The results are shown in Figure 5.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

```
Fixed-effects (within) regression               Number of obs     =        403
Group variable: province                        Number of groups  =         31

R-squared:                                      Obs per group:
     Within  = 0.9272                                         min =         13
     Between = 0.8869                                         avg =       13.0
     Overall = 0.8657                                         max =         13

                                                F(5,30)           =     294.03
corr(u_i, Xb) = -0.7894                         Prob > F          =     0.0000

                            (Std. err. adjusted for 31 clusters in province)
```

sales	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
GDP	.162544	.0296457	5.48	0.000	.1019993	.2230886
invest	.487594	.1036399	4.70	0.000	.2759331	.699255
area	.6605765	.0951945	6.94	0.000	.4661633	.8549897
expenditure	-.2327462	.0995622	-2.34	0.026	-.4360794	-.029413
construction	-.2432644	.0402156	-6.05	0.000	-.3253957	-.1611332
_cons	-2395.079	378.7773	-6.32	0.000	-3168.645	-1621.512
sigma_u	2173.7826					
sigma_e	614.25778					
rho	.92605546	(fraction of variance due to u_i)				

Figure 5. In-group estimator

The results in Figure 5 indicate that the constant term "cons" represents the average value of all individual effects. The notation 'rho=0.926' signifies that the variance of the composite disturbance term mainly stems from the variation of individual effects. It can be observed that the model passes the F-test. Additionally, at the 5% significance level, all variables are statistically significant. As an example, in the case of "area," holding other conditions constant, for every ten thousand square meters increase in the sales area of commercial housing, the average sales increases by 0.6606 hundred million yuan. The economic interpretations of the other explanatory variables are similar.

To further investigate whether to use a Mixed Effects Model or Fixed Effects Model, an F-test is performed using Stata software. The null hypothesis is "all ui equal to 0," which translates to "cannot reject the use of Mixed Effects Regression." The results of this test are presented in Figure 6.

```
Fixed-effects (within) regression               Number of obs     =        403
Group variable: province                        Number of groups  =         31

R-squared:                                      Obs per group:
     Within  = 0.9272                                         min =         13
     Between = 0.8869                                         avg =       13.0
     Overall = 0.8657                                         max =         13

                                                F(5,367)          =     935.08
corr(u_i, Xb) = -0.7894                         Prob > F          =     0.0000
```

sales	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
GDP	.162544	.0144155	11.28	0.000	.1341966	.1908914
invest	.487594	.0582023	8.38	0.000	.3731421	.6020459
area	.6605765	.0341675	19.33	0.000	.5933878	.7277651
expenditure	-.2327462	.055843	-4.17	0.000	-.3425586	-.1229338
construction	-.2432644	.0275077	-8.84	0.000	-.2973568	-.189172
_cons	-2395.079	115.3282	-20.77	0.000	-2621.866	-2168.292
sigma_u	2173.7826					
sigma_e	614.25778					
rho	.92605546	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(30, 367) = 23.34                    Prob > F = 0.0000
```

Figure 6. Feasibility test of fixed effect model

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

From the results in Figure 6, it is evident that the p-value of the F-test is extremely small. Therefore, we reject the null hypothesis, suggesting that the Fixed Effects (FE) model is superior to the Mixed Effects Regression. Thus, it is advisable to allow each individual to have their own intercept. However, as robust standard errors were not used in the analysis, further examination through the Least Squares Dummy Variable (LSDV) method is necessary. The LSDV method introduces individual fixed effects by incorporating specific effects of each individual as dummy variables into the regression model. This helps control for heterogeneity among individuals. These dummy variables can capture individual-specific influences, aiding in the explanation of inter-individual differences. Therefore, the LSDV method can control for individual heterogeneity and allow for the interpretation of individual fixed effects on the results. It is suitable for panel data where unobservable characteristics or individual-specific influences exist among individuals. The results of the LSDV analysis are presented in Figure 7.

sales	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
GDP	.162544	.0308336	5.27	0.000	.0995733	.2255146
invest	.487594	.1077927	4.52	0.000	.267452	.707736
area	.6605765	.0990089	6.67	0.000	.4583734	.8627796
expenditure	-.2327462	.1035516	-2.25	0.032	-.4442268	-.0212656
construction	-.2432644	.041827	-5.82	0.000	-.3286866	-.1578422
province						
2	-149.2311	148.213	-1.01	0.322	-451.9224	153.4603
3	-3972.788	516.7333	-7.69	0.000	-5028.099	-2917.478
4	-1259.259	200.5935	-6.28	0.000	-1668.926	-849.5926
5	-1823.711	326.6589	-5.58	0.000	-2490.837	-1156.584
6	-2905.282	444.6591	-6.53	0.000	-3813.397	-1997.167
7	-965.7942	227.6161	-4.24	0.000	-1430.648	-500.9401
8	-1229.746	295.8808	-4.16	0.000	-1834.015	-625.4764
9	-8.485788	213.4346	-0.04	0.969	-444.3774	427.4059
10	-4762.609	1222.479	-3.90	0.001	-7259.243	-2265.975
11	194.2739	646.2124	0.30	0.766	-1125.468	1514.016
12	-3987.833	660.2479	-6.04	0.000	-5336.239	-2639.427
13	-2311.525	505.1343	-4.58	0.000	-3343.147	-1279.903
14	-1374.468	404.1006	-3.40	0.002	-2199.751	-549.1842
15	-8072.814	1203.032	-6.71	0.000	-10529.73	-5615.894
16	-6457.022	938.5067	-6.88	0.000	-8373.708	-4540.336
17	-2579.094	502.8249	-5.13	0.000	-3606	-1552.189
18	-4142.844	631.1986	-6.56	0.000	-5431.923	-2853.764
19	-6751.676	1377.065	-4.90	0.000	-9564.017	-3939.335
20	-2246.414	396.0863	-5.67	0.000	-3055.33	-1437.498
21	-149.0448	209.9917	-0.71	0.483	-577.9051	279.8155
22	-1966.756	463.5457	-4.24	0.000	-2913.442	-1020.069
23	-3652.864	780.3844	-4.68	0.000	-5246.622	-2059.107
24	-1824.508	372.7089	-4.90	0.000	-2585.681	-1063.335
25	-1996.377	325.3435	-6.14	0.000	-2660.817	-1331.937
26	385.9872	201.278	1.92	0.065	-25.07725	797.0516
27	-1834.56	267.482	-6.86	0.000	-2380.831	-1288.289
28	-468.4137	220.8585	-2.12	0.042	-919.467	-17.36046
29	114.32	181.6042	0.63	0.534	-256.5654	485.2053
30	-274.2668	177.4851	-1.55	0.133	-636.7398	88.20627
31	-811.1026	263.0651	-3.08	0.004	-1348.353	-273.8519
_cons	-224.63	156.3342	-1.44	0.161	-543.907	94.64698

Figure 7. LSDV analysis results

From the results in Figure 7, it is evident that the p-values of most individual dummy variables are extremely small, indicating significance at the 5% level. Hence, we can reject the null

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

hypothesis that all individual dummy variables are equal to zero. This implies the presence of individual effects, and the use of mixed effects regression should be avoided.

In the analysis of the Fixed Effects (FE) model, considering time effects by utilizing two-way fixed effects for research can be beneficial. Now, year-specific dummy variables are defined for analysis. The "cum" in the dummy variables indicates the cumulative years represented by the variable. Cumulative year dummy variables can be used to analyze time trends with cumulative effects, such as studying the accumulated impact of a variable over the past few years. This approach helps control for long-term trends and cumulative effects on the time dimension, enabling a more accurate interpretation and prediction of changes in panel data.

Figure 8 presents the results of the Fixed Effects (FE) model considering time effects. The base year is set as 2010 and is therefore not included in the regression results. In Figure 8, it is observed that the coefficients for the time effects "Year" are all negative, and most yearly dummy variables are significant at the 5% level. With the exception of the expenditure, the coefficients of the other explanatory variables are significant at the 5% significance level.

```
Fixed-effects (within) regression               Number of obs     =        403
Group variable: province                        Number of groups  =         31

R-squared:                                      Obs per group:
     Within  = 0.9382                                         min =         13
     Between = 0.9031                                         avg =       13.0
     Overall = 0.8815                                         max =         13

                                                F(17,30)          =     822.63
corr(u_i, Xb) = -0.7984                         Prob > F          =     0.0000

                             (Std. err. adjusted for 31 clusters in province)
```

sales	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
GDP	.1647564	.0314363	5.24	0.000	.1005549	.2289579
invest	.4550688	.0961144	4.73	0.000	.2587769	.6513606
area	.5669278	.0975996	5.81	0.000	.3676028	.7662529
expenditure	-.1668216	.1070152	-1.56	0.130	-.3853758	.0517325
construction	-.2064615	.0408385	-5.06	0.000	-.2898648	-.1230582
year2	-264.0053	51.13939	-5.16	0.000	-368.4459	-159.5648
year3	-315.5265	78.28815	-4.03	0.000	-475.4122	-155.6408
year4	-398.7311	93.20353	-4.28	0.000	-589.0781	-208.3841
year5	-623.3827	145.9816	-4.27	0.000	-921.5169	-325.2485
year6	-526.6289	158.4497	-3.32	0.002	-850.2264	-203.0314
year7	-337.665	156.8631	-2.15	0.040	-658.0222	-17.30793
year8	-371.9082	167.4604	-2.22	0.034	-713.9079	-29.90855
year9	-256.2958	172.5798	-1.49	0.148	-608.7509	96.1592
year10	-304.9304	169.9023	-1.79	0.083	-651.9173	42.05646
year11	-82.23614	195.1963	-0.42	0.677	-480.8801	316.4078
year12	-475.7802	188.103	-2.53	0.017	-859.9379	-91.62259
year13	-1151.678	299.9236	-3.84	0.001	-1764.203	-539.152
_cons	-2090.364	347.5405	-6.01	0.000	-2800.136	-1380.592
sigma_u	2090.129					
sigma_e	575.73197					
rho	.92947664	(fraction of variance due to u_i)				

Figure 8. Fixed effect model considering time effect

The joint significance test for all yearly dummy variables is conducted next, and the results are presented in Figure 9. The results from Figure 9 indicate an extremely small p-value, rejecting the null hypothesis of no time effect at a 5% significance level. This suggests that time effects should be included in the model for analysis.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

```
( 1)  year2 = 0
( 2)  year3 = 0
( 3)  year4 = 0
( 4)  year5 = 0
( 5)  year6 = 0
( 6)  year7 = 0
( 7)  year8 = 0
( 8)  year9 = 0
( 9)  year10 = 0
(10)  year11 = 0
(11)  year12 = 0
(12)  year13 = 0

      F( 12,    30) =    4.62
           Prob > F =    0.0003
```

Figure 9. Results of joint significance test of annual dummy variables

3.3. Random Effects Model

To further analyze random effects, the LM test process and results are displayed in Figure 10. This test is employed in panel data analysis to examine individual effects. Rejecting this null hypothesis indicates the presence of a random disturbance term ui reflecting individual characteristics in the original model. Therefore, the mixed regression model should not be used, and a random effects model should be chosen instead.

```
Breusch and Pagan Lagrangian multiplier test for random effects

        sales[province,t] = Xb + u[province] + e[province,t]

        Estimated results:
                         |       Var     SD = sqrt(Var)
                ---------+-----------------------------
                   sales |   1.54e+07       3927.902
                       e |   377312.6       614.2578
                       u |   281101.3       530.1899

        Test: Var(u) = 0
                             chibar2(01) =   171.64
                          Prob > chibar2 =   0.0000
```

Figure 10. LM test result

```
Random-effects ML regression                    Number of obs     =        403
Group variable: province                        Number of groups  =         31

Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =         13
                                                              avg =       13.0
                                                              max =         13

                                                LR chi2(5)        =    1003.61
Log likelihood = -3233.8653                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
       sales | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         GDP |   .1434501   .0147033     9.76   0.000     .1146321     .172268
      invest |   .5262911   .0583162     9.02   0.000     .4119936    .6405887
        area |   .6168246   .0351375    17.55   0.000     .5479563    .6856929
 expenditure |   -.184488   .0557848    -3.31   0.001    -.2938241   -.0751518
construction |   -.214139   .0276418    -7.75   0.000     -.268316    -.159962
       _cons |      -2272   369.6295    -6.15   0.000     -2996.46   -1547.539
-------------+----------------------------------------------------------------
    /sigma_u |   1951.414   273.5829                      1482.563    2568.535
    /sigma_e |   612.4433   22.63972                      569.6396    658.4633
         rho |   .9103329   .0239847                      .8537211    .9487748
------------------------------------------------------------------------------
LR test of sigma_u=0: chibar2(01) = 243.21             Prob >= chibar2 = 0.000
```

Figure 11. MLE estimation results of random effects model

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

The test result indicates that the LM test rejects the null hypothesis of "no individual random effects" at a significance level of 5%. This suggests that the random effects model should be chosen over the mixed regression model. As a comparison, we can also perform Maximum Likelihood Estimation (MLE) for the random effects. The MLE results are presented in Figure 11.

The results in Figure 11 show that the coefficient of all explanatory variables were significant. For the tests, the extremely small p-value suggests that we reject the null hypothesis at a 5% significance level. This further supports the presence of individual random effects, indicating that the random effects model should be used instead of the mixed regression model.

3.4. Hausman Test

The basic idea of the Hausman test is to compare the estimated results between the fixed effects model and the random effects model to assess the significance of individual effects. The results of the Hausman test, as shown in Figure 12, are crucial for determining whether to choose the fixed effects model or the random effects model.

	—— Coefficients ——			
	(b) FE1	(B) RE1	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
GDP	.162544	.0569105	.1056335	.0117978
invest	.487594	.734564	-.24697	.0296899
area	.6605765	.3786682	.2819082	.0246388
expenditure	-.2327462	.0200683	-.2528145	.0383062
construction	-.2432644	-.0757487	-.1675158	.0218782
_cons	-2395.079	-1609.127	-785.9518	.

```
                  b = Consistent under H0 and Ha; obtained from xtreg.
   B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

    chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            = 132.76
Prob > chi2 = 0.0000
(V_b-V_B is not positive definite)
```

Figure 12. Hausman test results

The Hausman test typically assumes that individual effects are random, corresponding to the random effects model, hence the difference is expected to be close to zero. If the test statistic significantly deviates from zero, it implies rejection of the hypothesis that individual effects are random, suggesting that the fixed effects model may be more appropriate. Conversely, if it's not significant, it indicates that the assumption of random individual effects might be more appropriate. From the results of the Hausman test presented in Figure 12, it can be observed that the null hypothesis, which posits that ui is uncorrelated with xit and zi, is rejected with a very small p-value. Therefore, the fixed effects model is deemed more suitable than the random effects model.

Given that in traditional Hausman tests, the null hypothesis assumes that the random effects model is the most efficient, it's important to note that the data used in this study, as mentioned earlier, exhibits substantial discrepancies between cluster-robust standard errors and ordinary standard errors. Consequently, utilizing the conventional Hausman test may lead to biased results. To address this, a Hausman test employing cluster-robust standard errors is conducted. This is carried out using the 'xtoverid' command in Stata, and relies on the over-identification test based on the excess conditions imposed. This is because, compared to the fixed effects model, the random effects model introduces additional constraints, assuming that individual

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

heterogeneity (ui) is uncorrelated with the explanatory variables. Therefore, it can be viewed as an over-identification condition. The results of the Hausman test with robust standard errors are illustrated in Figure 13.

```
Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re  robust cluster(province)
Sargan-Hansen statistic  57.403  Chi-sq(5)    P-value = 0.0000
```

Figure 13. Hausman test results for robust standard error

From the results in Figure 13, the Chi-sq(5) statistic is 57.403, with a p-value much smaller than 0.01. Consequently, at a 1% significance level, the null hypothesis is rejected, indicating that the fixed effects model is preferred over the random effects model. To facilitate the comparison of coefficient estimates and standard errors obtained using various methods, the results are summarized and presented in Figure 14.

Variable	OLS	FE_robust	FE_TW	RE
GDP	.03184425	.16254397	.16475641	.05691049
	.02061989	.02964573	.03143628	.03393229
invest	.93380649	.48759401	.45506876	.73456404
	.15566872	.10363992	.09611444	.14183317
area	.06074787	.66057649	.56692781	.37866824
	.05748719	.09519453	.09759963	.07288942
expenditure	.03196987	-.23274619	-.16682161	.0200683
	.08094978	.09956224	.10701517	.08949743
construction	.00515662	-.24326443	-.20646149	-.07574866
	.04151953	.04021562	.04083846	.04122636
year2			-264.00532	
			51.139387	
year3			-315.52648	
			78.28815	
year4			-398.7311	
			93.203535	
year5			-623.38271	
			145.98161	
year6			-526.62892	
			158.44973	
year7			-337.66504	
			156.86307	
year8			-371.90823	
			167.46036	
year9			-256.29584	
			172.57983	
year10			-304.93041	
			169.90234	
year11			-82.236141	
			195.19625	
year12			-475.78024	
			188.10303	
year13			-1151.6777	
			299.92362	
_cons	-753.95282	-2395.0786	-2090.3639	-1609.1268
	201.30169	378.7773	347.54048	300.36932

Legend: b/se

Figure 14. Coefficient estimates and standard errors of each method

The results in Figure 14 demonstrate that OLS represents the mixed regression results; FE_robust stands for the fixed effects model using robust standard errors; FE_TW represents the fixed effects model considering time effects; and RE represents the random effects model results. It can be observed that there are significant differences in coefficient estimates among different methods. Overall, the fixed effects model exhibits substantial disparities compared to

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

other models. Additionally, the inclusion of time effects in the fixed effects model (FE_TW) has a certain impact on the estimated coefficients. As time effects have been demonstrated to be significant in the earlier analysis, this study chooses the fixed effects model with time effects (FE_TW) as the model for analyzing the panel dataset of the real estate industry.

In the FE_TW model, the explanatory variables, GDP, invest, area and construction significantly influence the dependent variable, while expenditure is not statistically significant. However, most coefficients for the explanatory variables are positive. The increase of regional Gross Domestic Product, real estate development investment and sales area of commercial housing will drive the sales of commercial housing to a certain extent. Besides, there is a certain degree of negative correlation between total output value of construction industry and sales of commercial housing.

4. Conclusion

This article conducted a modeling analysis using panel data comprising data on real estate sales and related influencing factors from various provinces and regions in China for the years 2010-2022. The panel data modeling process involved in-depth analysis using various models including mixed regression, fixed effects models, and random effects models. Additionally, robust standard errors and considerations of time effects in the fixed effects model were also investigated. Through model estimation, parameter estimation, and various tests such as LM test and Hausman test, it was ultimately determined that the fixed effects model with time effects (FE_TW) was most effective for analyzing the panel dataset of the real estate industry. In this article, it was found that regional Gross Domestic Product, the real estate development investment, the sales area of commercial housing and the total output value of construction industry significantly influence the sales of commercial housing in China.

Based on the above analysis, he following policy recommendations are proposed:

1) Considering the significant impact of real estate development investment on real estate sales, the government can implement measures to promote real estate development. This may include streamlining approval processes and providing loan support to attract more investment into the real estate market, thereby stimulating the growth of real estate sales.

2) Given the significant influence of real estate sales area on real estate sales, the government can encourage developers to increase the supply of real estate. This can be achieved by formulating corresponding policy measures to optimize land use planning and provide land development subsidies, among other strategies, to meet market demands and drive the growth of real estate sales.

3) The government can focus on enhancing residents' purchasing power. This can be achieved by improving employment levels, offering training opportunities, and enhancing the social security system, among other measures, to increase residents' income levels and stimulate housing demand.

4) The government can stabilize the real estate market through macroeconomic policies to avoid excessive investment and the formation of real estate bubbles. Specific measures may include establishing a sound regulatory mechanism for the real estate market and strengthening land supply management.

Acknowledgments

Minzu University of China.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0037

References

[1] Q. Tong. Empirical Analysis of the Relationship between Economic Growth and Environmental Pollution: Based on Panel Data from Nanjing City, 1986-2015 (China Collective Economy, China 2017), p.7-9. (In Chinese).

[2] J. Xia, R.Y. Liu. Digital Economy, Green Development, and Resource Allocation in the Tourism Industry: An Empirical Analysis Based on Provincial Panel Data in China (Guangxi Social Sciences, China 2023), No.334(04), p.129-138. (In Chinese).

[3] J.H. Fan. Study on the Relationship between Environmental Regulation, Technological Innovation, and Regional Economic Growth (MS., Tianjin Vocational Technical Normal University, China 2023), p.8.

[4] S.Z. Hua. Empirical Study on the Impact of Green Finance on the Ecological Economy of the Yellow River Basin (MS., Shanghai Normal University, China 2023), p.16.

[5] G.Y. Shen. Empirical Analysis of the Main Influencing Factors of Real Estate Prices: A Case Study of Xining City, Qinghai Province. (Qinghai Finance, China 2019), No.350(11), p.27-31. (In Chinese).

[6] J.H. Hu. Analysis of the Factors Affecting Real Estate Prices in China (MS., University of International Business and Economics, China 2019), p.6.

[7] J.M. Wu. Study on the Changes of Shanghai Housing Prices from 2003 to 2017 and Its Influencing Factors (Modern Commerce and Industry, China 2020), p.7-10. (In Chinese).

[8] L. Xu. Exploring Factors Affecting the Changes in National Average Housing Prices: An Analysis Based on Econometric Models (Vitality, China 2023) No.644(07), p.196-198. (In Chinese).

[9] Q. Chen. Advanced Econometrics and Its Applications with Stata (Higher Education Press, China 2014), p. 258-270.

[10] Information on http://www.stats.gov.cn.