The Impact of House Purchase Restriction to House Price in Beijing: A Regression Discontinuity Design Analysis

1. Background Introduction

China's economic growth is significantly driven by real estate. Around 2015, Beijing's house prices rose rapidly, leading to concerns of a potential bubble (Yan & Hongbing, 2018). To mitigate the growth, the government implemented continous policies from March 2017 (Lianjia, 2017), such as loan restrictions, to control house prices. The project analyzes the impact of these policies using causal inference based on the dataset of Beijing house prices.

Time	Policy
2017.03.17	The down payment for secondary homes increased to 60%, and corporate-owned properties had a 3-year resale restriction.
2017.03.21	The commercial loan interest rate for the first home decreased from 10% discount to 5%
2017.03.28	Real estate agents are barred from transacting the same property within a one-year period.
2017.04.03	Purchase restrictions are imposed on residential bungalows located on state-owned land.
2017.04.14	Commercial properties owned by enterprises can only be leased, not sold.
2017.05.01	The interest rate for first-home loans reverts to the benchmark rate, whereas the rate for second-home loans rises by 20%.

Table 1:Restriction Policies in Beijing’s Real Estate Market

2. Data Description

The dataset contains 318,851 housing transactions in Beijing from 2009 to 2018, sourced from Lianjia Network. Primarily focusing on 2011 to 2017, it shows house characteristics like location (latitude, longitude, district), transaction time, physical attributes (size, rooms, floors), and factors influencing price, like subway accessibility.

3. Research Question

Through the observation of Beijing's housing prices, it can be clearly found that from 2010 to 2017, Beijing's housing prices continued to rise, but in 2017 showed a clear downward trend (see figure 1).

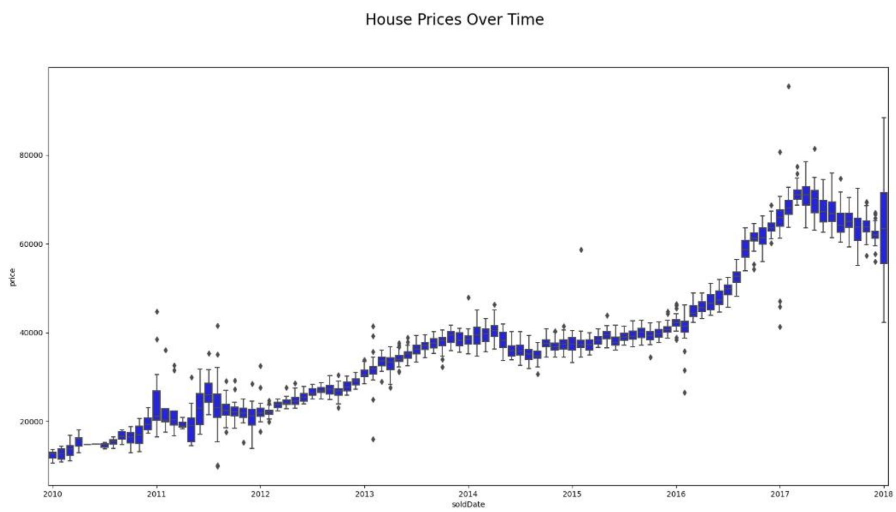

Figure 1: Bejing Housing Prices Trend

Considering the policy and the dataset's trends, examining the effects of restriction policies on housing prices in 2017 emerges as an interesting topic. Scholars (Sun, Zheng and Geltner, 2017) have conducted similar research using regression discontinuity design (RDD). According to their findings, house prices are decreased due to the restriction policy.

Extending the model of earlier studies, the project employs data from house price to study the restriction policies' impact on house price.

4. Data Cleaning and Variable Construction

Data cleaning and processing is crucial in empirical analysis as it ensures the availability of high-quality and well-structured data.

In the dataset, for null values in several columns, such as “DOM”, “buildingType”, we employ the mode () function to determine the most frequently occurring values within their respective columns. This method is particularly suitable for categorical variables, minimizing potential bias introduced by using average values.

Specifically, for the column “communityAverage”, we find it has a strong correlation with “Cid” and “district”, so we fill null values based on its grouped average on Cids and districts.

Furthermore, some cells have entries like '未知' in 'constructionTime' or '#NAME?' in 'livingRoom' which indicates unknown characters, we also use the same method to fill them to ensure consistency.

In addition, we restructure columns for data processing, we transfer ‘tradeTime’ into 'tradeYear', 'tradeMonth', and 'tradeDay' columns. This transformation makes the original data easier to analyze in further steps.

For better statistical analysis, logarithmic transformations are performed on columns with high skewness, and dummy variables are created for categorical data.

Figures 2 (geographic graph of Beijing) show that price per square declines spread outward with geographic location, underscoring the importance of location to house price valuations. We calculated the distance to major landmarks in Beijing, created the new feature Tiananmen Square, the center of Beijing, and used the

"Lag" and "Lat" columns of the dataset to explore their relationship with other columns.

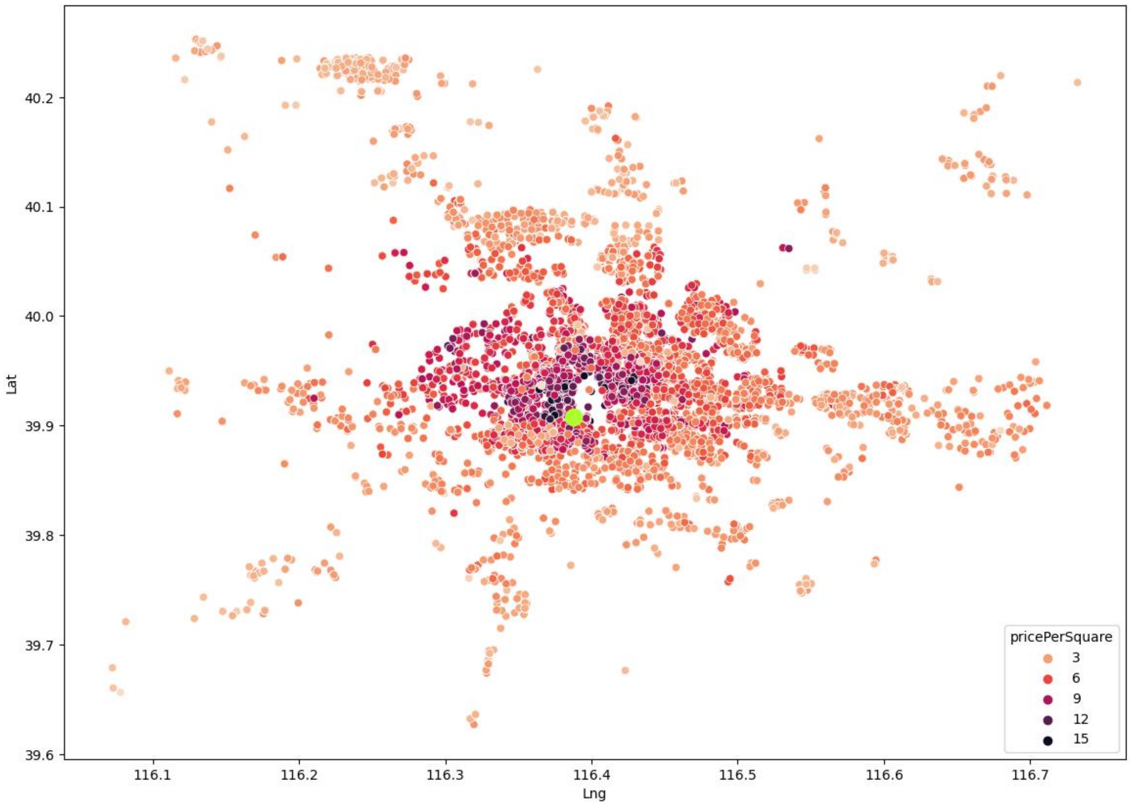

Figure 2: Beijing House Price with the Location of Tiananmen

5. Model Specification and Estimation

To examine the impact of restriction policies on housing prices, we choose regression discontinuity design (RDD) to evaluate. In contrast to Ordinary Least Squares, RDD is adept at managing the structural shifts across diverse time frames, as opposed to resorting to a constant linear model. The model is utilized separately with and without control variables, to ensure control over other factors affecting housing prices.

To choose the most suitable control variables, we can assess their correlations with the price and opt for variables with an absolute value greater than 0.5. Then through LASSO cross validation, we can apply the optimal alpha for LASSO in order to select our final control variables by penalty of AICc.

The model in this project is outlined as follows:

$$lnprice_i = \beta_0 + \beta_1 D_i + \beta_2\, Tradedays_i + \beta_3\, Tradedays_i\, D_i + \sum_{n=1}^{N} \alpha_n\, Z_{n,i} + \varepsilon_i$$

Here, $lnprice_i$ denotes the percentge change in price, β_1 represents the effect for the restriction policy D_i. $Tradedays_i$ is the number of days between the trade time of the sample point and the cutoff date. β_2 indicates the effect of $Tradedays_i$, while β_3 signifies the effect of policy D_i changes over the $Tradedays_i$. $Z_{n,i}$ represents the situation of each control variable in each individual, with α_n denoting the effect of each control variable. ε_i represent the unmeasurable white noise. (Li, Zhu, Zhao, Zheng and Zhang, 2020)

The cutoff for the RDD is set at March 17, 2017, the date when the government introduced the restriction policy. Acknowledging the potential lag between policy announcement and effect, the model establishes cutoffs with lag time of one month (April 17, 2017) and two months (May 17, 2017) .

6. Trends Exploration

Prior to analysis, we employ LOWESS to explore the trends within the existing variables for all three cutoffs. Notable jump effects around these cutoffs are observed, which may imply a significant impact of the policy. The graphs below show that post the policy's introduction, housing prices shifted from growth to a decline. This downtrend can be observed across all three cutoffs, highlighting the possible impact of the policy measures on housing prices.

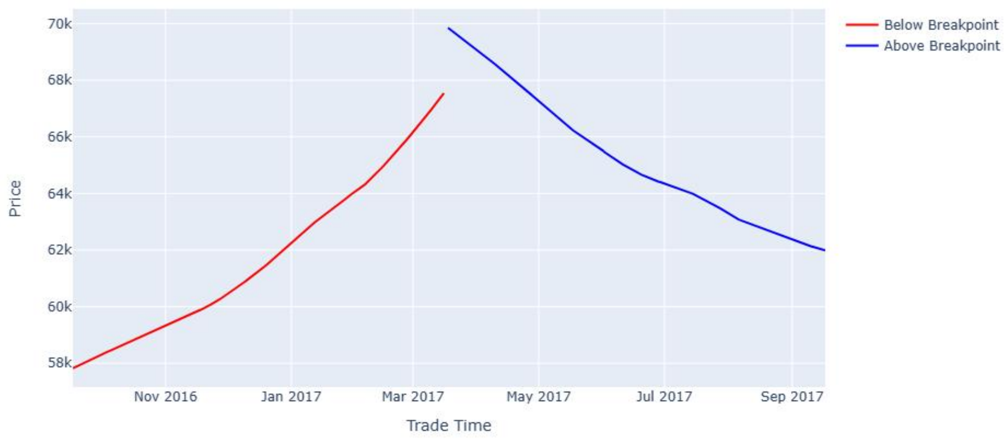

Figure 3: LOWESS for Cutoff on March 17, 2017

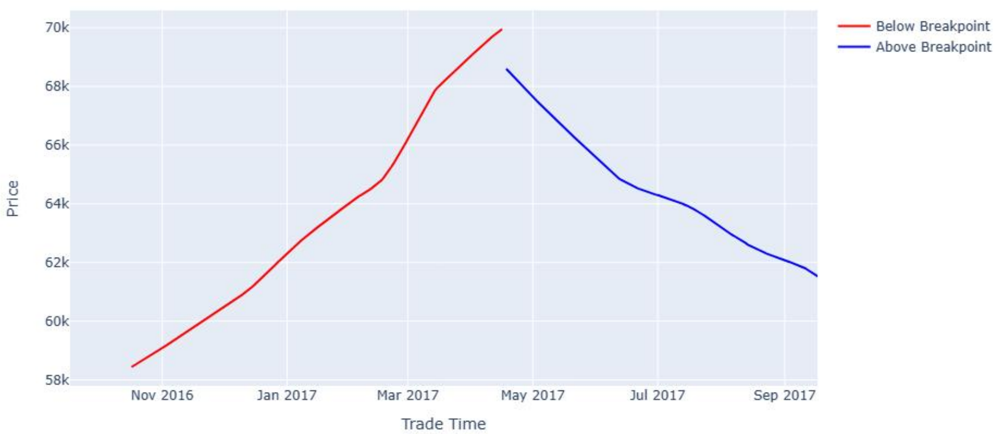

Figure 4: LOWESS for Cutoff on April 17, 2017

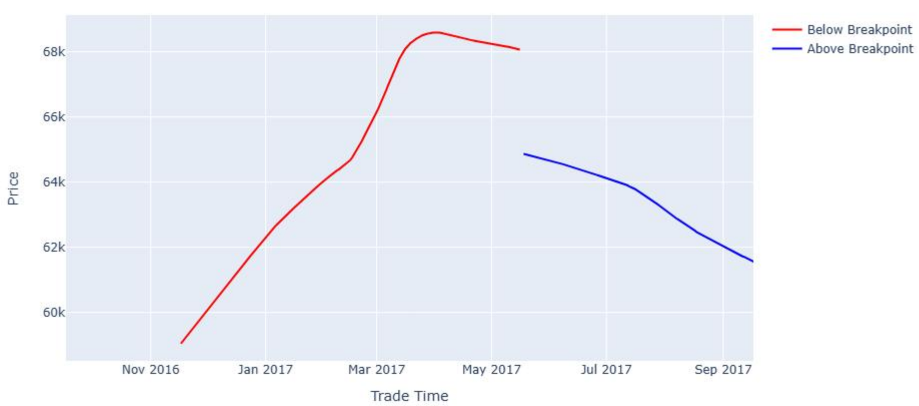

Figure 5: LOWESS for Cutoff on May 17, 2017

7. Empirical Result

Through the analysis of six models, we can infer the subtle effects of policies on house prices. In both Model 1 and Model 2, the positive impact on housing prices after the policy announcement is obvious. However, over time, this positive trend has significantly weakened. Model 3 shows that since the policy was introduced, its impact on housing prices has changed from positive correlation to negative correlation over time.

After adding control variables, only Model 4 shows a positive effect of the policy. In contrast, models 5 and 6 not only highlight significant negative effects but amplify the percentage effects. Moreover, a consistent theme across all six models is a significant shift over time from a positive to a diminishing correlation between policy and house prices.

	Model1	Model2	Model3	Model4	Model5	Model6
Policy effect	0.0096* (0.006)	0.0177*** (0.006)	-0.0634*** (0.006)	0.0255*** (0.002)	-0.0585*** (0.003)	-0.1131*** (0.003)
Interaction of policy and tradeday	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	No	No	No	Yes	Yes	Yes
Cutoff	2017.3.17	2017.4.17	2017.5.17	2017.3.17	2017.4.17	2017.5.17

Table 2: Effect of the House Price Restriction Policy on House Prices

8. Conclusion

In conclusion, through RDD, considering different time periods and whether control variables are introduced, we construct six models for in-depth analysis. The adoption of this methodology ensures the rigor of our research and provides a clear perspective on the complex effects of policy effects on house prices. In addition, our findings reveal a temporal evolution: an initial small increase in house prices following the release of the policy was replaced over time by a significant decline. The study highlights the inherent complexity and shifting nature of policy's impact on real estate dynamics, and the eventual decline in prices after policy implementation highlights the need for policymakers to take a long-term view to ensure the sustainability and effectiveness of such interventions.

Appendix

Table 3: Regression without Control Variables, Cutoff at 03.17.2017

Dep. Variable:	lnprice	**R-squared:**	0.081
Model:	OLS	**Adj. R-squared:**	0.081
Method:	Least Squares	**F-statistic:**	2342.
Date:	Fri, 20 Oct 2023	**Prob (F-statistic):**	0.00
Time:	07:14:30	**Log-Likelihood:**	-35294.
No. Observations:	79864	**AIC:**	7.060e+04
Df Residuals:	79860	**BIC:**	7.063e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	11.1182	0.003	3720.529	0.000	11.112	11.124
policy	0.0096	0.006	1.659	0.097	-0.002	0.021
tradedays	0.0014	1.95e-05	72.029	0.000	0.001	0.001
tradeday*policy	-0.0020	3.93e-05	-51.016	0.000	-0.002	-0.002

Omnibus:	9097.272	**Durbin-Watson:**	1.960
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	52754.404
Skew:	-0.397	**Prob(JB):**	0.00
Kurtosis:	6.902	**Cond. No.**	711.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 4: Regression without Control Variables, Cutoff at 04.17.2017

Dep. Variable:	lnprice	**R-squared:**	0.051
Model:	OLS	**Adj. R-squared:**	0.051
Method:	Least Squares	**F-statistic:**	1303.
Date:	Fri, 20 Oct 2023	**Prob (F-statistic):**	0.00
Time:	07:14:44	**Log-Likelihood:**	-31132.
No. Observations:	72865	**AIC:**	6.227e+04
Df Residuals:	72861	**BIC:**	6.231e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	11.1073	0.003	3616.585	0.000	11.101	11.113
policy	0.0177	0.006	3.154	0.002	0.007	0.029
tradedays	0.0012	2.32e-05	53.455	0.000	0.001	0.001
tradeday*policy	-0.0018	3.65e-05	-49.503	0.000	-0.002	-0.002

Omnibus:	7799.099	**Durbin-Watson:**	1.961
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	41597.808
Skew:	-0.387	**Prob(JB):**	0.00
Kurtosis:	6.620	**Cond. No.**	689.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified

Table 5: Regression without Control Variables, Cutoff at 05.17.2017

Dep. Variable:	lnprice	**R-squared:**	0.024
Model:	OLS	**Adj. R-squared:**	0.024
Method:	Least Squares	**F-statistic:**	503.0
Date:	Fri, 20 Oct 2023	**Prob (F-statistic):**	5.43e-323
Time:	07:14:55	**Log-Likelihood:**	-25609.
No. Observations:	62356	**AIC:**	5.123e+04
Df Residuals:	62352	**BIC:**	5.126e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	11.1538	0.005	2241.371	0.000	11.144	11.164
policy	-0.0634	0.006	-10.650	0.000	-0.075	-0.052
tradedays	0.0010	3.23e-05	30.628	0.000	0.001	0.001
tradeday*policy	-0.0016	4.1e-05	-37.863	0.000	-0.002	-0.001

Omnibus:	3984.001	**Durbin-Watson:**	1.963
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	13453.459
Skew:	-0.277	**Prob(JB):**	0.00
Kurtosis:	5.207	**Cond. No.**	784.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 6: Regression with Control Variables, Cutoff at 03.17.2017

Dep. Variable:	lnprice	**R-squared (uncentered):**	1.000
Model:	OLS	**Adj. R-squared (uncentered):**	1.000
Method:	Least Squares	**F-statistic:**	3.975e+07
Date:	Fri, 20 Oct 2023	**Prob (F-statistic):**	0.00
Time:	07:21:05	**Log-Likelihood:**	39125.
No. Observations:	79864	**AIC:**	-7.823e+04
Df Residuals:	79853	**BIC:**	-7.812e+04
Df Model:	11		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
tradedays*policy	-0.0023	1.55e-05	-147.333	0.000	-0.002	-0.002
square	-0.0883	0.001	-66.636	0.000	-0.091	-0.086
tradeYear	0.0004	1.51e-05	29.167	0.000	0.000	0.000
disTianAnMen	-0.1501	0.010	-14.929	0.000	-0.170	-0.130
district_6	-0.0177	0.002	-10.361	0.000	-0.021	-0.014
district_10	-0.0088	0.002	-4.001	0.000	-0.013	-0.004
subway_1.0	0.0013	0.001	1.091	0.275	-0.001	0.004
district_8	0.0037	0.002	1.989	0.047	5.44e-05	0.007
policy	0.0255	0.002	11.110	0.000	0.021	0.030
tradedays	0.0015	7.69e-06	197.788	0.000	0.002	0.002
communityAverage	0.9690	0.003	364.136	0.000	0.964	0.974

Omnibus:	107990.481	**Durbin-Watson:**	1.917
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	138798599.530
Skew:	-7.123	**Prob(JB):**	0.00
Kurtosis:	206.734	**Cond. No.**	3.91e+04

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large, 3.91e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Table 7: Regression with Control Variables, Cutoff at 04.17.2017

Dep. Variable:	lnprice	**R-squared (uncentered):**	1.000
Model:	OLS	**Adj. R-squared (uncentered):**	1.000
Method:	Least Squares	**F-statistic:**	3.787e+07
Date:	Fri, 20 Oct 2023	**Prob (F-statistic):**	0.00
Time:	07:22:08	**Log-Likelihood:**	37084.
No. Observations:	72865	**AIC:**	-7.415e+04
Df Residuals:	72854	**BIC:**	-7.404e+04
Df Model:	11		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
tradedays*policy	-0.0020	1.72e-05	-117.626	0.000	-0.002	-0.002
square	-0.0905	0.001	-66.489	0.000	-0.093	-0.088
tradeYear	0.0005	1.55e-05	34.029	0.000	0.000	0.001
disTianAnMen	-0.1514	0.010	-14.697	0.000	-0.172	-0.131
district_6	-0.0191	0.002	-10.891	0.000	-0.023	-0.016
district_10	-0.0093	0.002	-4.119	0.000	-0.014	-0.005
subway_1.0	0.0005	0.001	0.422	0.673	-0.002	0.003
district_8	-0.0011	0.002	-0.572	0.567	-0.005	0.003
policy	-0.0585	0.003	-22.092	0.000	-0.064	-0.053
tradedays	0.0014	8.28e-06	172.360	0.000	0.001	0.001
communityAverage	0.9575	0.003	349.612	0.000	0.952	0.963

Omnibus:	96673.246	**Durbin-Watson:**	1.919
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	108475986.510
Skew:	-6.897	**Prob(JB):**	0.00
Kurtosis:	191.518	**Cond. No.**	3.90e+04

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large, 3.9e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Table 8: Regression with Control Variables, Cutoff at 05.17.2017

Dep. Variable:	lnprice	**R-squared (uncentered):**	1.000
Model:	OLS	**Adj. R-squared (uncentered):**	1.000
Method:	Least Squares	**F-statistic:**	3.578e+07
Date:	Fri, 20 Oct 2023	**Prob (F-statistic):**	0.00
Time:	07:22:39	**Log-Likelihood:**	34654.
No. Observations:	62356	**AIC:**	-6.929e+04
Df Residuals:	62345	**BIC:**	-6.919e+04
Df Model:	11		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
tradedays*policy	-0.0017	1.85e-05	-90.062	0.000	-0.002	-0.002
square	-0.0908	0.001	-64.439	0.000	-0.094	-0.088
tradeYear	0.0006	1.61e-05	37.210	0.000	0.001	0.001
disTianAnMen	-0.1485	0.011	-13.998	0.000	-0.169	-0.128
district_6	-0.0226	0.002	-12.563	0.000	-0.026	-0.019
district_10	-0.0060	0.002	-2.555	0.011	-0.011	-0.001
subway_1.0	-8.502e-05	0.001	-0.068	0.946	-0.003	0.002
district_8	-0.0028	0.002	-1.427	0.154	-0.007	0.001
policy	-0.1131	0.003	-42.398	0.000	-0.118	-0.108
tradedays	0.0012	1.03e-05	116.535	0.000	0.001	0.001
communityAverage	0.9462	0.003	333.346	0.000	0.941	0.952

Omnibus:	74023.418	**Durbin-Watson:**	1.929
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	45063407.564
Skew:	-5.711	**Prob(JB):**	0.00
Kurtosis:	134.202	**Cond. No.**	3.90e+04

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large, 3.9e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Reference

1. Li,Yaoyao & Zhu, Daolin & Zhao, Jiangmeng & Zheng, Xiaodong & Zhang, Lanyue, 2020. Effect of the housing purchase restriction policy on the Real Estate Market: Evidence from a typical suburb of Beijing, China, Land Use Policy, Elsevier, vol. 94(C).

2. Sun, W., Zheng, S., Geltner, D.M., Wang, R., 2017. The housing market effects of local home purchase restrictions: evidence from Beijing. J. Real Estate Financ. Econ. 55, 288–312.

3. Yan,Yan & Hongbing, Ouyang, 2018, Effects of house-sale restrictions in China: a difference-in-difference approach, Applied Economics Letters, 25:15, 1051-1057

4. Lianjia. (2017, December 25). 2017 Beijing Real Estate Market Policy Highlights: What Has Affected Your Purchase of a House This Year? Zhihu.

5. https://www.kaggle.com/code/ericliu123/beijing-houseprice-predict

Python Code Link:

https://drive.google.com/file/d/1W2DnOQH7hkpuXVlKWTZyCqrL3aJ5UZ0e/view?usp=sharing