Online platformazation and Offline expansion of Real Estate Brokerage: Evidence from China[*]

Version Zero [†]   Jan 2024[‡]

Jan 2024
Preliminary and Incomplete.

Abstract

This is the main result of the baseline regression and dynamic analysis. You can read the Model and Results section to get more details. Moreover, the Discussion and Conclusion section are the most important part, because this is what we should discuss to make the paper better.

[†]University of Michigan. zxuyuan@umich.edu
[‡]Feel Free to modify

The remainder of the paper proceeds as follows. Section 4 describes the model. We then present our empirical results in Section 5. Finally, Section 6 provides the discussion and Section 7 concludes.

1 Introduction

Brief introduction and summary of papers.

2 Background

Brief background.

3 Data

This section describes the data and statistical summary. Should consist of graph and one big table.

4 Model

This is the main result that we have currently done.

The dependent variable is *the natural logarithm of lianjia's income*, *the natural logarithm of lianjia's transaction price* and *the natural logarithm of negotiation period* and in order to quantify the effect of lianjia, we propose three sets of explanatory variables. First, we construct a density based index (DBI) to capture the effect of lianjia's effect and continuous expansion effect. It is defined as the Second, we construct a proxy variable to capture the effect of the continuous DBI effect. Third, we construct a set of variables to capture the effect of lianjia's entry and continuous expansion.

There are three sets of explanatory variables, including the density-based index (DBI), entry and its continuous effects, and proxy variables capturing the continuous effects of the DBI. First, the DBI is defined as the number of lianjia's stores in a given region as a percentage of the total number of stores in that region. The index is a direct measure of the market power of lianjia in that region. Second, the entry effect of lianjia and its continuous effect are defined as a set of dummy variables consisting of the time of lianjia's entry into that particular market and the corresponding pre- and post-periods. Since we only have seven years in our panel data, we construct seven periods as: $Pre_3 Pre_2, Pre_1, Entry, Post_1, Post_2, Post_3$, while in our regression model, we drop Pre_1 in the model and use it as the reference group. The proxy variable is defined as the product of lianjia's market entry and its continuous effects and the DBI, which captures lianjia's local market power. We do not combine all indicators in a single regression because they may interact with each other, making our estimation and interpolation difficult.

We first construct the multi-way clustering with fixed effect model to estimate DBI on the dependent variables. The model is as follows:

$$Y_{it} = \beta_0 + \beta_1 density_{it} + \alpha \mathbf{X_{it}} + \tau_{\mathbf{it}} + \mu_{\mathbf{i}} + \epsilon_{\mathbf{it}}. \tag{1}$$

where Y_{it} is the three main dependent variables, including $\log(income), \log(price)$, and $\log(period)$, $density_{it}$ is the DBI, $\mathbf{X_{it}}$ is a set of control variables, including brokerage_control L_hedonic_control, transaction_control and region_control, while τ_{it} is the time dummy variable interacting with the fixed effect of the business area, μ_i is the individual fixed effect and ϵ_{it} is the error term. The standard errors are clustered at the individual level.

When it comes to the estimation of the entry effect and its continuous effect, we use the following model:

$$
\begin{aligned}
Y_{it} = {} & \beta_0 + \beta_1 Pre_3 + \beta_2 Pre_2 + \beta_3 Entry \\
& + \beta_4 Post_1 + \beta_5 Post_2 + \beta_6 Post_3 + \alpha \mathbf{X_{it}} + \tau_{\mathbf{it}} + \mu_{\mathbf{i}} + \epsilon_{\mathbf{it}}.
\end{aligned} \tag{2}
$$

where Pre is the pre-period of lianjia's entry, $Entry$ is the entry period, and $Post$ is the post-period of lianjia's entry. All other model settings are the same as the previous model.

Finally, we construct a proxy variable to capture the continuous effect of the DBI. The model is as follows:

$$
\begin{aligned}
Y_{it} = {} & \beta_0 + \beta_1 proxy\_entry + \beta_2 proxy\_pos\_1 \\
& + \beta_3 proxy\_pos\_2 + \beta_4 proxy\_pos\_3 + \alpha \mathbf{X_{it}} + \tau_{\mathbf{it}} + \mu_{\mathbf{i}} + \epsilon_{\mathbf{it}}.
\end{aligned} \tag{3}
$$

where $proxy\_entry$ is the product of the entry effect and the DBI, and $proxy\_pos\_i$ is the product of the post-period effect and the DBI. All other model settings are the same as the previous model. The results are shown in Appendix Table A1 and Appendix Table A2.

However, our estimates may not be asymptotically efficient due to the endogeneity of DBI and serial correlation issues. First, DBI is endogenous because the entry of lianjia is endogenous. This is due to the fact that lianjia will choose better locations with higher potential revenues in order to maximize profits. Second, the model suffers from serial correlation because our sample is during the boom period of China's real estate market, and as the number of transactions in the real estate market increases, so do the revenues and prices in the brokerage industry. In addition, lianjia tend to be concentrated in popular areas, which makes for greater growth and thus confounds our estimates. To solve this problem, we use a dynamic panel model for estimation. Dynamic panel models can capture the dynamics of the process by using the lagged dependent variable as a regression variable. This helps in understanding how the past values of the variables affect their current values.

To estimate the dynamic panel model, we use the Arellano-Bond (1991) GMM estimator, which

is a two-step estimator. The first step is to use the lagged dependent variable as an instrument for the current dependent variable, and the second step is to use the lagged residuals as an instrument for the current dependent variable. In this paper, we do not use system GMM estimator because the system GMM must assume that the lagged explanatory variables are independent of factors that do not vary over time in the region, and this clearly cannot hold in the present setting. The model is as follows: [**this should be a line that describes what measures that we use to construct the model**]

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \beta_1 \Delta Measure_{it} + \alpha \mathbf{\Delta X_{it}} + \mathbf{\Delta \tau_{it}} + \mathbf{\Delta \epsilon_{it}}. \tag{4}$$

where ΔY_{it} is the first difference of the dependent variable, $\Delta Measure_{it}$ is the first difference of the our three measurement methods, $\mathbf{\Delta X_{it}}$ is the first difference of the control variables, $\Delta \tau_{it}$ is the first difference of the time dummy variable interacting with the fixed effect of the business area, and $\Delta \epsilon_{it}$ is the first difference of the error term. The standard errors are clustered at the individual level. The results are shown in Appendix Table A3.

5 Results

5.1 Main Results

For Table A1, the result reveals a significant correlation between the lianjia's local market power and lianjia's income in this given community. However, we discovered no substantial influence of lianjia's local market power on the housing prices or transaction periods in the same community. This could be interpreted to mean that whilst Lianjia's local market power could induce an increase in income, it does not necessarily grant them any significant pricing power nor expedite the transaction period. Therefore, we may deduce that the augmentation of Lianjia's income is primarily driven by an increase in the number of transactions, presumably due to more property listings by Lianjia within the community. This suggests that although lianjia can control more market share in the local market, it does not necessarily mean that it can have monopoly power in this given region. This is partially because the price of the house is determined by the market, and there are also some potential other brokerages that can enter the market. However, lianjia's local market power can still have a significant influence on the income because they can list more housings and have more transactions in the local market.

To further test the results, we first check if lianjia's entry effect and continuous effect is significant. For column 1, 3, and 5 of the Table A2, the result reveals that before the lianjia's entry, the natural logarithm of income, housing price and transaction period are not significant. After the entry, we see that after lianjia's entry into this local market, lianjia can immediately increase their income and

keep making significantly more income comparing with previous periods. Besides, we see that the entry do not significantly increase the housing price immediately but significantly increase housing price one year after but later on turning back to the normal pricing. Lastly, the lianjia's entry do not significantly lowers the transaction periods. This is consistent with our previous results that lianjia's entry can significantly increase the income but not the housing price and transaction periods in equilibrium.

Lastly, we use the third measurement to test after the lianjia's entry, the continuous effect of lianjia's local power of the local market. For column 2, 4, and 6 of the Table A2, the result reveals that the continuous effect of lianjia's local power of the local market is significant for the income, housing price and transaction periods. This result indicates that in different period, the lianjia's local market power can have different influence on the market.

To deal with the endogeneity problem, I propose to use the GMM method to estimate the model. [**this should be a line that describes the method we construct the GMM model**]

The results in Table A3 show that the statistical significance of DBI is robust but our OLS model tends to underestimate the result. Furthermore, we can see that in different period, the lianjia's local market power can make lianjia gain significant more income in the long run but not in the short run after adjusting the serial correlation problems. Besides, lianjia's local market power can also significantly lower the housing prices in this community as well as significantly lower the transaction periods in the long run. This result means that our previous fixed effect model may not be asymptotically efficient and we shall use the GMM method to estimate the model.

Here, is the most important, we should carry out our causal inference and validate our GMM result! We should work on this right now! This is what we should discuss January 2.

5.2 Robustness Check

This section is pretty easy to construct, we can do various tricks. Moreover, we can select different regions with high and low night time lights to find out the different influence factors of lianjia. We just leave it here.

6 Discussion

Pool or separate is also a key issue that we need to take into consideration. Here are some potential ways that we can make the paper better. This exogenous shock can make sure our specification is correct, and we can discuss the following methods.

1. Think about the RD design. There may be potential sharp discontinuties in the geosystem. We shall think about it geographically.

2. Think about the DID design. Because we have multiple cities and multiple years, we can select some cities, which have policies in these years, and treat them as the treatment group and other cities as the control group.

3. Think about potential IV. This is pretty hard -_-.

7   Conclusion

Overall, this is the current results. Just finish the baseline regression and some dynamic analysis.

References

Appendix Table A1. Baseline Regression Results

	(1) log(income)	(2) log(price)	(3) log(period)
density	0.0726***	0.00374	-0.0184
	(0.0264)	(0.00427)	(0.0264)
lianjia_5	-0.0000826	0.000234	0.00278
	(0.00413)	(0.000816)	(0.00424)
other_5	0.00205***	0.000890***	-0.00111
	(0.000784)	(0.000141)	(0.000755)
ln_lead	0.0974***	0.00819***	0.197***
	(0.00476)	(0.000872)	(0.00553)
ln_watch_people	0.0239***	-0.0148***	0.212***
	(0.00373)	(0.000703)	(0.00444)
ln_negotiation_period	0.0418***	-0.00683***	
	(0.00371)	(0.000787)	
ln_watch_time	0.0264***	-0.000145	0.0399***
	(0.00214)	(0.000342)	(0.00248)
ln_nego_changes	0.00297	-0.0102***	0.631***
	(0.00601)	(0.00134)	(0.00635)
N	134648	134648	134648
R-squared	0.885	0.989	0.762

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table A2. Dynamic Regression Results

	(1) log(income)	(2) log(income)	(3) log(price)	(4) log(price)	(5) log(period)	(6) log(period)
pre3	-0.00976 (0.0105)		0.0000211 (0.00182)		0.0163 (0.0108)	
pre2	0.00604 (0.00802)		-0.00212 (0.00134)		0.00690 (0.00859)	
entry	0.0339*** (0.00818)		0.00120 (0.00159)		0.00236 (0.00832)	
post1	0.0192** (0.00797)		0.00235* (0.00141)		-0.000120 (0.00823)	
post2	0.00689 (0.00809)		0.000315 (0.00164)		-0.00453 (0.00814)	
post3	0.0234*** (0.00843)		-0.000136 (0.00131)		-0.0128 (0.00873)	
lianjia_5	-0.00553 (0.00456)	-0.00212 (0.00451)	-0.000217 (0.000998)	0.000109 (0.000969)	0.00168 (0.00465)	0.00357 (0.00463)
other_5	0.00153** (0.000752)	0.00186** (0.000768)	0.000861*** (0.000135)	0.000886*** (0.000138)	-0.00108 (0.000722)	-0.00129* (0.000741)
ln_lead	0.0972*** (0.00476)	0.0973*** (0.00476)	0.00818*** (0.000872)	0.00819*** (0.000872)	0.224*** (0.00537)	0.224*** (0.00537)
ln_watch_people	0.0239*** (0.00373)	0.0239*** (0.00373)	-0.0148*** (0.000704)	-0.0148*** (0.000705)	0.213*** (0.00441)	0.213*** (0.00441)
non_online_effect	-0.157*** (0.0187)	-0.156*** (0.0187)	0.00414 (0.00402)	0.00414 (0.00402)	0.447*** (0.0294)	0.447*** (0.0294)
ln_negotiation_period	0.0418*** (0.00371)	0.0418*** (0.00371)	-0.00683*** (0.000788)	-0.00683*** (0.000788)		
ln_watch_time	0.0265*** (0.00214)	0.0264*** (0.00214)	-0.000143 (0.000342)	-0.000145 (0.000343)	0.0356*** (0.00249)	0.0356*** (0.00249)
ln_nego_changes	0.00313 (0.00601)	0.00308 (0.00601)	-0.0102*** (0.00134)	-0.0102*** (0.00134)	0.624*** (0.00633)	0.624*** (0.00633)
proxy_entry		0.0853*** (0.0224)		0.00209 (0.00412)		-0.00443 (0.0222)
proxy_pos1		0.0400 (0.0253)		0.00723* (0.00437)		-0.0185 (0.0264)
proxy_pos2		-0.0234 (0.0287)		-0.000159 (0.00616)		-0.0544* (0.0297)
proxy_pos3		0.0222 (0.0331)		0.00478 (0.00482)		-0.0720** (0.0336)
N	134648	134648	134648	134648	134648	134648
R-squared	0.885	0.885	0.989	0.989	0.763	0.763

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table A3. GMM Regression Results

	(1) log(income)	(2) log(income)	(3) log(price)	(4) log(price)	(5) log(period)	(6) log(period)
L.ln_income	0.0634*** (0.00567)	0.0657*** (0.00538)				
density	0.875*** (0.116)		-0.0192 (0.0127)		0.0873 (0.385)	
lianjia_5	-0.236*** (0.0298)	-0.00372 (0.00470)	0.00648** (0.00277)	0.00668** (0.00287)	-0.0332 (0.0762)	0.193*** (0.0441)
other_5	0.0165*** (0.00242)	-0.00146** (0.000677)	-0.000557 (0.000936)	-0.000440 (0.000875)	-0.00570 (0.0246)	-0.0470*** (0.0116)
ln_lead	0.0853*** (0.0117)	0.0543*** (0.00475)	0.00797*** (0.000950)	0.00795*** (0.000951)	0.355*** (0.0295)	0.291*** (0.0172)
ln_watch_people	0.00179 (0.00504)	0.0111*** (0.00401)	-0.0163*** (0.000841)	-0.0163*** (0.000844)	0.289*** (0.0390)	0.157*** (0.0134)
ln_negotiation_period	0.0673*** (0.00336)	0.0690*** (0.00362)	-0.000369 (0.000777)	-0.000317 (0.000780)		
ln_watch_time	0.00658*** (0.00199)	0.00640* (0.00333)	-0.00377*** (0.000470)	-0.00379*** (0.000469)	0.0460*** (0.0141)	0.0963*** (0.0112)
ln_nego_changes	-0.0208*** (0.00622)	-0.00824 (0.0110)	-0.0161*** (0.00179)	-0.0160*** (0.00179)	0.844*** (0.0685)	0.600*** (0.0199)
proxy_entry		0.0292 (0.0224)		-0.0106 (0.00860)		-0.679*** (0.156)
proxy_pos1		0.0255 (0.0257)		-0.0166* (0.00914)		-0.342** (0.138)
proxy_pos2		0.0304 (0.0287)		-0.0326*** (0.00958)		-0.112 (0.129)
proxy_pos3		0.0753** (0.0339)		-0.0147* (0.00857)		-0.268** (0.132)
L.ln_end_price			-0.0421*** (0.0116)	-0.0456*** (0.0110)		
L.ln_negotiation_period					1.091*** (0.273)	0.0353 (0.0243)
L2.ln_negotiation_period					-0.281*** (0.0790)	-0.0394*** (0.0145)
observations	97589	97589	97589	97589	64429	64429
AR(1) p-value	0	0	2.34e-16	1.75e-16	0.000521	6.20e-09
AR(2) p-value	0.476	0.306	0.182	0.131	0.136	0.210
Hansen J Test p-value	0.165	0.180	0.341	0.328	0.177	0.224

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$