Instant Happiness or Regret?

The Impact of iBuyers in Real Estate Market Liquidity

Larasati Wulandari (1004751262)

Renato Mateus Magela Zimmermann (1004767978)

University of Toronto Department of Economics

April 14, 2021

Abstract

EDIT

1 Introduction and Motivation

Land is and has been one of the strongest storages of value available for modern humans. It has been the driving force behind countless wars and more recently financial innovations and catastrophes. The importance of land, and more specifically real estate, has been especially heightened recently, with the results of the 2008 subprime mortgage crisis, and currently stands as a haven for investors and individuals seeking stability. Still, one issue that has limited the real estate market for years is its lack of liquidity, which affects especially individuals whose purchase marks a significant portion of their wealth.

With the objective of tackling this issue, and backed by massive amounts of capital and data, Instant Buying (commonly referred to as IBuying) has grown to become a popular method of investing in residential real estate. As their name suggests, IBuyers differ from traditional residential real estate investors as they reduce the time from offer to purchase to often less than a week. IBuyers leverage price data and statistical models to evaluate properties at that speed, such that they can buy the property at a discount and hopefully resell it at competitive prices within a short amount of time.

Since the success of Opendoor, the first major IBuyer, in 2014, several other firms have joined the market for quick buying and selling of residential real estate. While most firms operate at a local level, others, such as real estate listing giant Zillow, have set up operations throughout the United States and brought with them billions of dollars in venture capital. While these companies appeal to individuals for their quick, all-cash offers, they often try advertising themselves as generators of higher market liquidity and fairer prices.

At their best, IBuyers back their claims with annecdotal evidence rather than numbers, and at the worst, IBuers could have a detrimental effect on markets. Though companies tout their qualities as home buyers, the true effect on a region's real estate market depends on their qualities as sellers. Given enough market share in a given area, IBuyers can create a local monopoly if they own a significant portion of available homes. If this is indeed an IBuyer's ultimate strategy, we should be seeing higher prices and a possible increase in the time needed for a property to be sold.

In this study, we aim to test this claim that IBuyers improve market liquidity and affect market prices in severl US counties. We will explore whether the IBuyer's claim is in fact true, and whether their presence in a county can in fact affect property prices and reduce the time needed to sell a house in their acting region.

2 Literature Review

3 Institutional Background

4 Data

The data used in the analysis was gathered from various sources through official sources and using web scraping techniques. The Subsection **??** discusses the general structure of our final dataset. Section 4.2 breaks down the data sources and important features, as well as any significant flaw or limitation they might have. Finally, Subsection 4.3 outlines key statistics of our selected data.

4.1 Structure

The full dataset has a panel format, and contains 121 characteristics related to the economics, demographics and real estate market of 798 different US counties for the years of

2016 and 2019. Of these counties, 112 are identified as treatment (having one or more active iBuyer in 2019). Of the aforementioned 149 characteristics, 23 are selected as our regressors; a detailed discussion of this selection is provided in Section 5.

The years present in the dataset are constrained to 2013 and 2016 due to limitations in our data and unnecessary variation in the year 2020. The Number of iBuyers dataset is composed of data scraped from the web in February 2021. Due to the lack of historical data available in our source, we are forced to limit our analysis to include only pre and post treatment data.

The year 2019 was selected as our post-treatment sample date to avoid any noise from the COVID-19 pandemic; several economic characteristics included in our analysis where significantly impacted by the pandemic, which could introduce unnecessary unobserved variation in our model .This is further discussed in **??** and in Section 5. The year 2016 was selected as

The year 2016 was selected as our pre-treatment date to conform to limitations in housing market data mentioned in IV. Note that the optimal pre-treatment date would be 2013, which is one year before *Opendoor*, the first modern iBuyer, was founded. The three years in between that and our data did not see much nationwide growth in the industry, with only one other major iBuyer, *Offerpad*, being founded in 2015. The year 2016 is still well-set as the pre-treatment year of the data, as it is one year before two large players, *RedfinNow* and *Orchard*, where founded and two years before the online real estate *Zillow* began operating in the iBuying market.

4.2 Composition

The full dataset is the result of merging seven datasets. This section briefly outlines the origin of said datasets, the variables selected from each of them and their key limitations. For a discussion on the data processing and methods used to collect and merge each dataset, refer to Subsection 9.1 in the Appendix.

I Number of iBuyers

The Number of iBuyers dataset (referred to as n_ibuyers in the accompanying code) is composed of data scraped from the website *ibuyer.com*[**?**] in February 15, 2021. The code for this collection accompanies this paper. The complete dataset contains the number of local and top iBuyers in a given county upon collection (which we label 2019 for consistency with remaining data).

The information gathered from the website is limited to the number of "local" or "top" iBuyers in each US City surveyed by the website. Top iBuyers are a group of iBuyer companies with significant investments towards iBuying who are active throughout the United States. This list is composed of *Opendoor*, *RedfinNow*, *Orchard*, *Offerpad* and *Zillow Offers*, which where selected by the website in accordance to these standards. Of these, we do not include cities in which *Opendoor* or *Offerpad* where active, as these two companies where founded before 2016. Smaller iBuyers must register with the website to be included as a “Local iBuyer”, which in turn constrains our treatment group to counties with at least one IBuyer registered in the website.

We follow with the assumption that events in 2020 did not significantly affect the number of counties with active iBuyers compared to 2019. This is a fair assumption, as by 2020 the iBuying market was already active in most major counties, and with the COVID-19 pandemic, there is no reason to believe iBuyers significantly expanded their market share. We consider this issue further in Section 5, where we compare the effects of including the number of active iBuyers, in our model, compared to a dummy variable stating iBuyers where present in a county.

One additional consideration is that the data available on the website is in the city level, while our analysis is on the county level. This further increases the need to treat the presence of iBuyers as a dummy variable rather than an ordinal discrete data point. This need stems from how any omitted city would not translate into an omitted county, but rather bias the number of iBuyers in a county downwards.

II GDP

This dataset is composed of various characteristics of production in a given county and year. The raw data has panel format and is subset to include only the years 2016 and 2019.

III Population Information

This dataset is composed of aggregate population statistics in a given county and year. The raw data has panel format and is subset to include only the years 2016 and 2019.

IV Housing

This dataset is composed of real estate listing information in a given county and month. The raw data has panel format and is subset to include only the years 2016 and 2019. This dataset has data ranging as far back as 2016 and is the reason for us selecting 2016 to be our pre-treatment year. This is an especially important dataset as it contains our regressand, the median days on market variable.

The data is originally monthly, and so we aggregate numeric columns, such that the final data shows the values for the month of December. The decision of taking the last month instead of an average is based on the objective of keeping values as close to the date the number of iBuyers was scraped as possible. It is worth mentioning that no variable in this dataset was cummulative, which does not warrant for using summation.

V Education

This dataset is composed of achievement scores in English and Mathematics for public schools in a given county. The raw data has cross-sectional format, so the final data is the result of merging data from 2016 and 2019.

For both Mathematics and English, we retain the percentage of students in a county who score above a ”proficient” level. This percentage is often a range to account for measurement uncertainties in a given county. The data present in our final dataset is the mean of said range.

VI Wages

This dataset is composed of wage and employment statistics in a given county and year. The raw data has panel format and is subset to include only the years 2016 and 2019.

VII Characteristics

This is composed of four datasets containing **Social**, **Demographic**, **Economic** and **Housing** characteristics for a given county and year. Each of these datasets has panel format and is subset to include only the years 2016 and 2019.

These four datasets contains 13 of the 23 selected characteristics.

VIII Others

Other datasets are used for the purpose of merging the datasets discussed above. A more detailed treatment of this process is described in Section 9.1 of the Appendix. These datasets include:

- A table of information on all US cities used to merge counties and cities and area FIPS code.

- A map of US Counties and their respective LEAID codes.

4.3 Summary Statistics

I General Statistics

The statistics referred to here are available in Table 3 in the Appendix. The first thing to notice in our selected dataset is the high range of standard deviations present in the selected variables. Regressors, such as the median listing price, median household income, population size estimate and construction spending have standard deviation values above five figures while variables such as homeowner vacancy rate and median rooms, which have standard deviations as low as 0.53. The same observation applies to the means of the data. This is a natural result of the various units of the various variables, and warrants normalizing the data. This normalization is further dicussed in Section 5.

We further highlight the mean of our regressands. As seen in Table 3, the mean median listing price amongst counties is 274354.04. This is to be expected in the US housing market, and follows what is expected to be seen in home sales. The high mean is accompanied by a significant variance of 162236.86 which illustrates the fairly broad range of listing prices in the market. Still, as we can see in Figure 1 this data is highly skewed to the right, following the shape of an exponential distribution.

The mean median days on market in turn has a fairly different statistical appearance compared to the median listing price. The mean median days on market is 78.55 (about two months and a half), and has standard deviation of 23.45. This is to be expected from an average marken in the United States. As seen in Figure 2, the data follows a relatively normal distribution, though it also presents a slight rightward skew.

II Treatment/Control Comparative Statistics

The statistics referred to here are aviailable in Table 4. This table compares the mean of each selected variable for control, treatment and both values. Both of the regressands, median listing price and median days on market, are fairly different. The treatment group has an average median listing price 81539.85 dollars (31%) higher than the control group. The treatment group also has an average median days on market 11.59 days below (-14.5%) that of the control group. These two characteristics are apparent from Figures 3 and 4 in the Appendix.

Other values that significantly differ when comparing the control and treatment groups are the population estimate (267214.26 vs 897316.43), rate of natural income (2.26 vs. 4.22), average household size (14475.26 vs. 48343.23) and construction spending (510128.97 vs. 1937689.42), where comarissons are for control group versus treatment group averages. These differences are consistent with a more urban profile for the treatment group compared to the control group. This makes economic sense, as larger, more urban markets are more likely to yield higher returns to real estate investors such as iBuyers. It also corresponds to areas with higher digital literacy and available data.

5 Empirical Framework

6 Results

7 Interpretation

8 Conclusion

Table 1: Dataset of Origin of Selected Regressors

regressor_name	dataset
median_rooms	characteristics
educ_nohs	characteristics
white	characteristics
mean_cash_public_assistance_income_dollars	characteristics
per_capita_income_dollars	characteristics
median_household_income_dollars	characteristics
part of educ_further	characteristics
mean_retirement_income_dollars	characteristics
median_age_years	characteristics
mean_travel_time_to_work_minutes	characteristics
homeowner_vacancy_rate	characteristics
move_post2010	characteristics
average_household_size	characteristics
english_prof_pct	educ
math_prof_pct	educ
construction	gdp
edian_listing_price	h_vals
median_square_feet	h_vals
median_days_on_market	h_vals
nlocal_ibs	n_ibuyers
ntop_ibs	n_ibuyers
rnaturalinc	pop_info
rnetmig	pop_info
popestimate	pop_info
annual_avg_emplvl	wages

9 Appendix

9.1 Data Processing and Methods

9.2 Charts

9.3 Figures

Table 2: General Summary Statistics

	mean	std	min	max
median_listing_price	274354.042614	1.622369e+05	67500.000000	2.020050e+06
median_days_on_market	78.553030	2.345368e+01	23.000000	1.780000e+02
median_rooms	5.693813	5.329772e-01	3.300000	8.000000e+00
homeowner_vacancy_rate	1.639657	1.087029e+00	0.000000	7.800000e+00
mean_travel_time_to_work_minutes	25.214646	5.247091e+00	14.400000	4.710000e+01
math_prof_pct	46.819991	1.281202e+01	11.444444	9.000000e+01
english_prof_pct	51.249663	1.155455e+01	15.416667	8.637500e+01
median_square_feet	1827.165404	5.912998e+02	0.000000	4.463000e+03
median_age_years	38.617677	4.654999e+00	24.400000	6.740000e+01
moved_in_2010_to_2014	19.848801	7.973362e+00	6.700000	3.860000e+01
annual_avg_emplvl	154349.486427	2.943917e+05	8782.000000	4.509905e+06
median_household_income_dollars	61671.342487	1.647529e+04	31207.000000	1.518000e+05
per_capita_income_dollars	31669.158775	7.945113e+03	13359.000000	8.272000e+04
popestimate	356319.613952	7.038671e+05	61473.000000	1.063444e+07
rnaturalinc	2.533129	3.524344e+00	-11.786826	1.737025e+01
rnetmig	4.029436	1.067996e+01	-65.400546	5.385107e+01
less_than_9th_grade	5.481124	3.200229e+00	0.400000	1.960000e+01
associates_degree	13.664110	6.255577e+00	3.300000	3.700000e+01
average_household_size	19264.671730	4.752123e+04	2.120000	9.890780e+05
white	42.600758	4.056286e+01	0.100000	9.880000e+01
mean_cash_public_assistance_income_dollars	2772.248037	1.138819e+03	269.000000	1.082400e+04
construction	712006.203283	1.410527e+06	0.000000	1.864062e+07
mean_retirement_income_dollars	26170.630997	5.861282e+03	12944.000000	5.589500e+04

Table 3: General Summary Statistics

	mean	std
treatment	0.141414	3.485583e-01
median_listing_price	274354.042614	1.622369e+05
median_days_on_market	78.553030	2.345368e+01
median_rooms	5.693813	5.329772e-01
homeowner_vacancy_rate	1.639657	1.087029e+00
mean_travel_time_to_work_minutes	25.214646	5.247091e+00
math_prof_pct	46.819991	1.281202e+01
english_prof_pct	51.249663	1.155455e+01
median_square_feet	1827.165404	5.912998e+02
median_age_years	38.617677	4.654999e+00
moved_in_2010_to_2014	19.848801	7.973362e+00
annual_avg_emplvl	154349.486427	2.943917e+05
median_household_income_dollars	61671.342487	1.647529e+04
per_capita_income_dollars	31669.158775	7.945113e+03
popestimate	356319.613952	7.038671e+05
rnaturalinc	2.533129	3.524344e+00
rnetmig	4.029436	1.067996e+01
less_than_9th_grade	5.481124	3.200229e+00
associates_degree	13.664110	6.255577e+00
average_household_size	19264.671730	4.752123e+04
white	42.600758	4.056286e+01
mean_cash_public_assistance_income_dollars	2772.248037	1.138819e+03
construction	712006.203283	1.410527e+06
mean_retirement_income_dollars	26170.630997	5.861282e+03

Table 4: Treatment/Control Summary Statistics

treatment	0	1	All
median_listing_price	262823.15	344363.01	274354.04
median_days_on_market	80.19	68.60	78.55
median_rooms	5.73	5.45	5.69
homeowner_vacancy_rate	1.64	1.61	1.64
mean_travel_time_to_work_minutes	25.22	25.20	25.21
math_prof_pct	47.03	45.53	46.82
english_prof_pct	51.57	49.29	51.25
median_square_feet	1812.60	1915.57	1827.17
median_age_years	38.92	36.75	38.62
moved_in_2010_to_2014	19.62	21.23	19.85
annual_avg_emplvl	110745.13	419090.23	154349.49
median_household_income_dollars	61437.53	63090.91	61671.34
per_capita_income_dollars	31415.57	33208.79	31669.16
popestimate	267214.26	897316.43	356319.61
rnaturalinc	2.26	4.22	2.53
rnetmig	4.05	3.88	4.03
less_than_9th_grade	5.39	6.01	5.48
associates_degree	13.51	14.60	13.66
average_household_size	14475.26	48343.23	19264.67
white	43.19	39.04	42.60
mean_cash_public_assistance_income_dollars	2720.58	3049.60	2772.25
construction	510128.97	1937689.42	712006.20
mean_retirement_income_dollars	25906.35	27775.20	26170.63

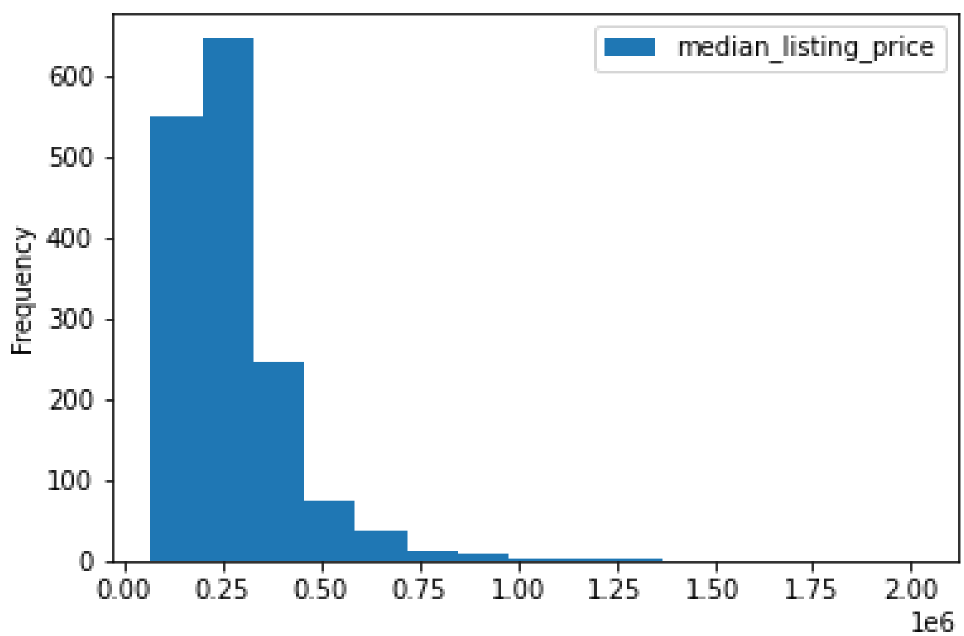

Figure 1: Histogram for Median Listing Price
(Aggregate of 2016 and 2019)

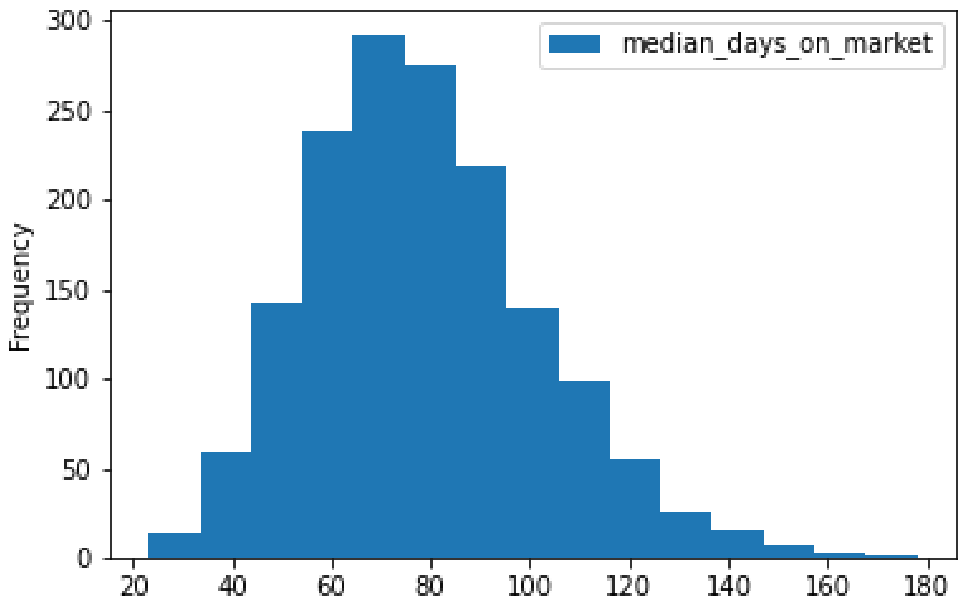

Figure 2: Histogram for Median Days on Market
(Aggregate of 2016 and 2019)

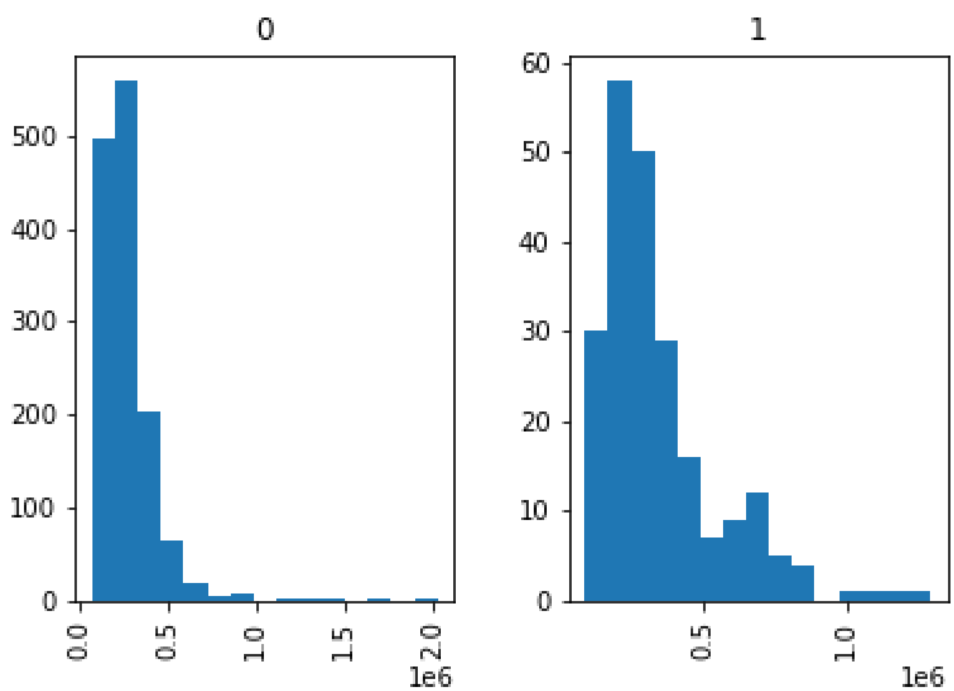

Figure 3: Control/Treatment Comparisson for Median Listing Price

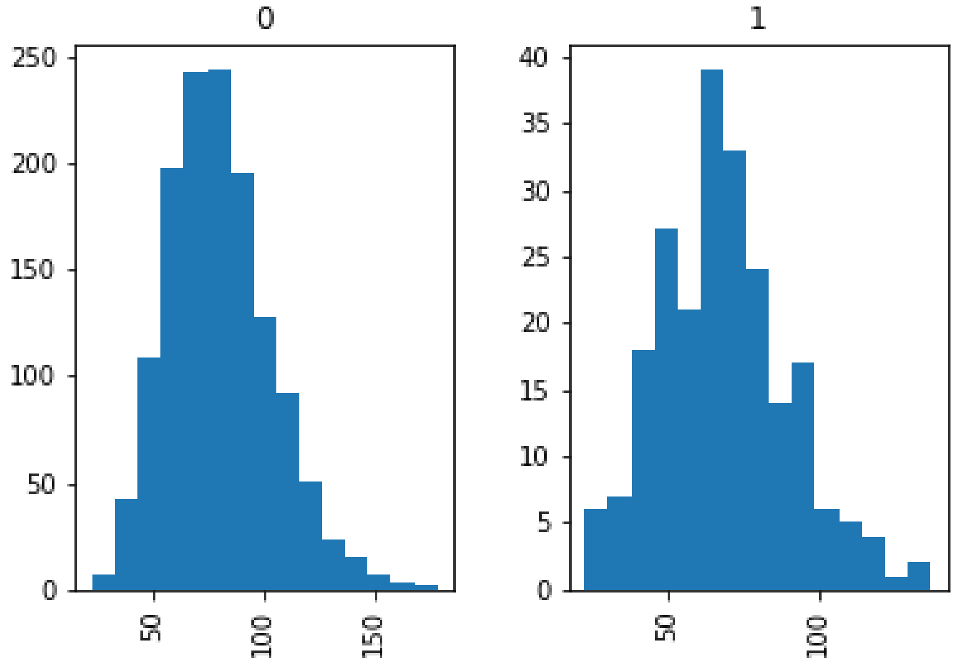

Figure 4: Control/Treatment Comparisson for Median Days on Market