The Effect of Market-Rate Housing Development on Surrounding Housing Prices: A Preliminary Study

Kerry Cook, Nripesh Kumar, Sotiris Karagounis
Columbia University

Abstract

Unaffordable housing costs has become a large issue in American cities and is a topic that is increasingly debated. There has been discussion on how new market rate developments affect housing costs in surrounding areas. In this paper, we outline our process to study the effects of newly completed market-rate developments on rent prices of nearby apartments in New York City. We found that there is a trend between new developments and higher rental prices in the surrounding areas. The methods outlined in this paper can be used to further understand the dynamics between new developments and surround housing costs.

1 Introduction

Housing in several modern American urban areas over the past decades has become increasingly expensive. For households in cities with large job growth, housing is often the largest expense in comparison to other needs, such as transportation or food. Stagnation in wages over time as one of the culprits of the increasing difficulty to pay for housing (Anacker, 2019). Large increases in housing prices acutely affect renters and could make homeownership more difficult to obtain, while current homeowners may see an increase in wealth. It should be noted that this phenomenon is not unique to American urban areas, in fact many cities both in Europe, as well as in Asia are dealing with their own housing affordability crises (Anacker, 2019).

Solutions to decrease housing costs have been the topic of much debate. Discussions involve housing supply and whether increasing the supply of market rate housing would decrease housing prices in urban areas. Economists argue that housing is inherently an economic issue that adheres to the laws of supply and demand. Due to large job growth in urban areas such as San Francisco, New York, and Boston, pro-developers argue that demand has outpaced supply in these cities. By increasing the supply of housing through new developments at market-rate prices, the larger supply could alleviate the demand which puts upwards pressure on the price of housing. Research performed by Asquith, Mast, and Reed (2019) show that while there is a correlation between new construction and rising rents, this effect is observed since developers select neighborhoods that are already undergoing changes (Asquith et al., 2019). Supporting the economists, they conclude that if new housing is built, higher income houses will choose this option instead of displacing residents from already existing units. Supply skeptics argue for the introduction of more regulation, such as rent controls and zoning laws, with a stronger alignment towards new subsidized developments at affordable housing rates. One argument for supply skepticism is that long time residents of a neighborhood may be displaced by new developments at market-rates, as their neighborhoods may become unaffordable. Several studies found that there may be a "spill-over" effect on prices in the immediate neighborhood of a new development depending on various local and property specific variables (Zahirovich-Herbert and Gibler, 2014; Ooi and Le, 2013). Understanding how prices change around a new development could lead to a better understanding of where to place new market-rate developments so as to not displace local residents and to ensure equity in access to housing.

2 Hypothesis

A research question of interest is: *How do new market-rate developments affect housing prices of neighborhoods in New York?* This study looks to test the following hypothesis:

- *New York rental prices rise when a new market-rate development is constructed in close proximity.* Prior research has mainly used housing prices and property evaluations to estimate the effects of new market-rate developments. Research on the rental market is more sparse; however, it could be a more insightful area of research on the effect of new developments on surrounding neighborhoods, as rentals switch tenants quite often and hence may be more sensitive to changes in the housing market.

3 Prior Work

An analysis of supply skepticism (Been et al., 2019) attempted to address some of the supply skeptic's arguments. The paper makes the distinction that supply skeptics can be homeowners or affordable housing advocates. Homeowners worry that a new development would devalue their homes as an investment and would change the character of their neighborhood. Affordable housing advocates are not in favor of building market rate housing, as it causes the aforementioned displacement in lower income neighborhoods due to increased rents and cost of living as a spill-over effect. Ellen (2001) investigates the spill over effect of affordable housing development and found a positive increase in surrounding prices. Using the same analysis structure of Ellen (2001), several studies have attempted to observe the effects of market-rate developments (Ooi and Le, 2013; Zahirovich-Herbert and Gibler, 2014). Both papers conclude, given certain environmental variables, houses in the surrounding areas may observe a small increase in prices.

Ellen (2001) was one of the first papers to observe the spill-over effects from new developments. Specifically, this paper focusing on new subsidized affordable housing in the New York City area and its localized effects on housing prices. The method used to choose which houses to observe prices is through considering a ring around a new development with a certain radius, which in this case is from 500 ft to 2000 ft. Additionally, in order to estimate the effect, the paper models the house prices through hedonic regression. This type of regression uses housing prices as an output and controls for various various property and location specific variables, for example proximity to a new development and income level. The estimates of the price change is deemed to be significant if the difference in prices before and after a new development of houses within the defined ring is larger than the difference of prices of a control group outside the ring, which would help control for inflation. Using this method, the study concluded that new affordable housing projects increased the price of surroundings houses compared to the rest of the zip-code in which the observations were made (Ellen et al., 2001).

Ooi and Le (2013) investigated the phenomenon of the price "spill-over" effect from 275 new developments on under-utilized lots of land in Singapore. The paper hypothesizes that there are two opposing pressures on housing prices surrounding a new development. There is a downwards pressure on prices, as a new development may increase the supply of available units and thus the increased competition drives down prices. The upwards pressure on price, the paper argues, is an "amenity effect", where the new development may improve the aesthetics of the surrounding area, as well as, draw more high income earners to the neighborhood. In terms of the methodology, Ooi and Le (2013) is similar to Ellen (2001), where a 500m to 1km radius around a new development was used to record surrounding house prices. The housing prices were modeled using a hedonic regression model. Ooi and Le (2013) concludes that there is a significant, but small increase in prices in the surrounding homes after an in-fill development. The issue that arises with the analysis that the paper performs is that it assumes that new developments are randomly placed; however, this an unrealistic assumption given that market-rate developers are more likely to develop in neighborhoods that are becoming popular with high income individuals. Additionally, infill developments use plots of lands that are under-utilized or empty and hence the increase in prices seen in this study may be overstated for new developments on already developed land, where the amenity effect may be less drastic.

Zahirovich-Herbert and Gibler (2014) similarly attempted to estimate the effect of new market-rate housing in Baton Rouge on the surrounding housing prices during a two decade period. The paper differs from the Ooi and Le (2013) study, as it attempts to control for buildings that are "atypical" in size compared to the size of the surrounding housing stock. An example of such a atypical building could be an apartment complex in a suburb. The relative size of new developments is important, as it may change the aesthetic of the neighborhood and may lead to a decrease in surrounding house prices. These types of developments are especially worrisome for supply skeptic homeowners. Zahirovich-Herbert and Gibler (2014) uses geographic rings with a radius of a mile. Hedonic regression was used to control for various environmental variables and the paper performed a fixed effect analysis to control for any unobserved variables that may be important to price changes. The results of the study point to the level of spill over effect depends on the size of the new development, where a larger single family home than the neighborhood average leads to slight increase in neighborhood prices, while a house that adheres to the neighborhood sizes leads to a slight decrease.

A more recent study has attempted to measure the "spill over" effect on rentals rather than housing prices (Li, 2019). Li (2019) focused on the New York City area using the same method of constructing rings as in Ellen (2001). This paper differs in methodology from the others as it attempts to address the selection bias of new developments occurring in areas with already rising prices by controlling for prices during the time when the new development was approved. Interestingly, Li (2019) found that when 10% of housing stock is added, rent prices decrease by 1% in the surrounding area. The paper uses a smaller ring size of 500ft than those in the papers previously mentioned. As a result, the smaller ring may allow for a more direct estimation of the spill over effect, which may be the reason why a negative effect on price is observed. The paper argues that the supply effect is stronger than the amenity effect of a new development and hence is the reason why a negative effect on rental prices is observed.

Several papers may validate the arguments of supply skeptics that new developments will increase housing prices locally, which will lead to the displacement of long-term residents (Ooi and Le, 2013; Zahirovich-Herbert and Gibler, 2014; Ellen et al., 2001). Been (2019) argues that while local prices may increase, the additional supply of new housing will cause prices in other neighborhoods to fall. However, measuring this phenomenon may not be needed as studying the immediate area surrounding a new development may be more insightful. Li (2019) has shown that if proximity is close enough to a new development, the hyper-localized supply effect may be stronger than the amenity effect of new-developments (Li, 2019). Additionally, Been (2019) points out that it is difficult to truly measure the spill over effect, as neighborhoods that developers choose are likely to have already been rising in prices and hence the above studies on market-rate housing will have selection bias. Lastly, Been (2019) argues that studies such as Ooi and Le (2013) and Zahirovich-Herbert and Gibler (2014) use property values instead of rent, which may not be an accurate representation of housing costs as they are estimates which may be less or more receptive to new housing developments in the local area.

4 Method

The rental price of an apartment in New York City is determined by multiple attributes of the apartment, including square feet, neighborhood, and building amenities. We utilized a similar method to Ellen (2001), who used a hedonic regression in combination with the difference-in-differences technique to estimate the effect of home sale prices in Singapore in surrounding areas after the completion of a new in-fill development.

The difference-in-differences method is used to estimate the effect of a treatment (or a specific outside intervention) by comparing the changes to the desired dependent variable overtime between the treatment group and a control group. More specifically, the difference in differences is the difference in the dependent variable between the treatment group and the control before and after the treatment was introduced. The difference-in-difference method is particularly useful when the population is not randomized - the treatment introduced is correlated to the dependent variable. This method ensures that any outside factors that are not observed in the data that are correlated to the treatment decision and the dependent variable do not bias the estimated effect (Buckley and Shang, 2002). Similar to Ooi and Le (2013), we assume that developments in New York City are not randomly placed or selected, and that developers strategically select locations, such as up and coming neighborhoods. To control for this bias, we selected the difference in difference technique to estimate the effect of a new building project on surrounding rental prices.

In addition to handling potential selection bias in areas with new developments, we controlled for other factors that would cause an apartment to have a high rent price. We included variables to represent size and building amenities to control for higher prices due to more bedrooms and larger apartment, in addition to apartments with certain amenities, such as a doorman or a gym, that would add upward pressure to the rental price. Eq. (1) represents the final model used to measure the effect of a new building development on rental price:

$$R_{it} = \alpha T_i + \beta PROX_i + \theta PROXT_{it} + \gamma \mathbf{A_i} + \epsilon (1)$$

where R_{it} is the rental price of apartment i at time t, T_i is a binary variable that equals 1 if the rental price represents a price after a new development completion, $PROX_i$ is a binary variable equal to 1 if the apartment is located close to a new development, $PROXT_{it}$ is a product variable that equals 1 when apartment i is close to a new development and the rental price is after a new development completion, $\mathbf{A_i}$ represents a matrix of apartment features used as control. The matrix includes bedrooms, bathrooms, a binary flag to indicate amenities available, and binary variables to indicate which borough the apartment is located in (Manhattan, Queens, Brooklyn, or the Bronx). The estimated effect of a new development on surrounding rental prices from our model is represented by θ.

5 Data

The data used for this analysis were gathered from multiple public sources. Specifically the NYC Open Data platform was used to gather candidate buildings. The candidate building's attributes were then webscraped from Streeteasy, a popular rental site specific to the New York City rental market.

5.1 Gathering Candidate Buildings

The Housing New York Units by Building [1] dataset, provided by the the department of Housing Preservation and Development (HPD), was used to identify which buildings were new developments. This dataset contains buildings completed after 2014 and provides information on the housing developments, as well as, their corresponding completion dates. Specifically, the buildings listed in the HPD dataset are developments that contribute to the Housing New York plan introduced in 2014 (Glen, 2014). The plan is meant to spur the creation of new affordable housing through private developers. The incentives that are given to private developers are tax credits or Mandatory Inclusionary Housing through the up-zoning of land parcels (Glen, 2014). In essence, developers are given incentives to create a percentage of affordable housing in their market-rate buildings.

[1] Housing New York by Building Dataset; Department of Housing Preservation and Development (HPD)

Since, the HPD dataset contains the building's completion dates, a specific year could be chosen in order to study the prices of surrounding market-rate rentals before and after the completion date. For this study, only buildings completed in 2018 were considered. In the HPD dataset there are a total of 364 new developments completed in 2018. The treatment set for the quasi-experiment constructed in this study are buildings in proximity to new developments completed in 2018. The control group has been defined as buildings that are not in proximity to any new developments in the HPD dataset. Thus, the control group contains all buildings not in proximity to any new developments completed after 2014. If all buildings except those in proximity to developments completed in 2018 were included in the control group, there could be the possibility that confounding variables may be introduced between the treatment and the control set. In essence, this larger control group could be in proximity to another new development completed in a non-2018 year and could still be seeing the effects on price from buildings completed in other years. It should be noted that 2018 was the year chosen, as many StreetEasy price histories begin in 2016. This limiting factor only allowed for years after 2017 to be completion years for this study. 2019 could not be selected, as the prices after 2019 were affected by the COVID-19 pandemic and would not be appropriate for studying post-completion prices in surrounding rentals.

The choice of using the completion date of a new development as the benchmark for pre and post treatment rather than approval date is due to the fact that rentals differ from home-ownership, as immediate environmental factors are more important than price speculation. Specifically, if there is a construction site near a rental, it would be expected that the rental price would fall due to noise pollution and other negative externalities that arise from a construction site. Apartments for sale on the other hand may rise in price with an approval of a new development, as buyers may speculate that the amenity effect of the new building may increase the value of properties in the surrounding area. Hence, the completion year was chosen so as to observe the competing supply and amenity effect of a new development on rental prices.

To formalize the definition of "proximity" for this study, a similar technique was used as specified in both Ooi and Le (2013) and Zahirovich-Herbert

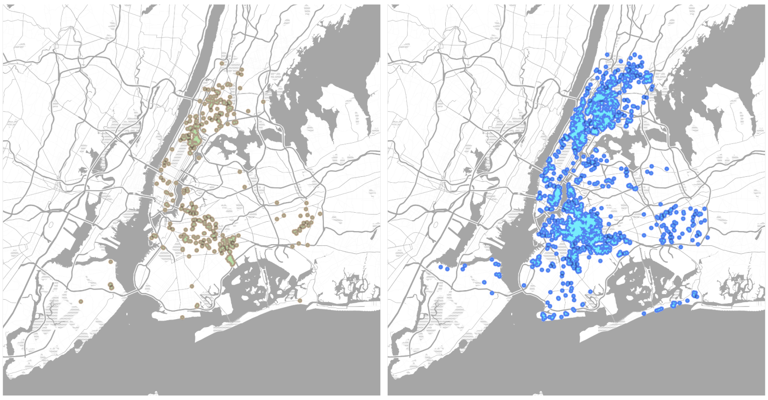

Figure 1: Above are geospatial representation of the area for which multiple dwelling buildings were separated into “in proximity” and “not in proximity to a new development”. The left figure shows all 250m radii around new developments specifically for the new developments that were completed in 2018. The buildings that belong in the radii defined will be “in proximity" to new developments. The right figure represents the area for which a building is within a 250m radius from any new developments in the HPD table after 2014. The buildings that fall outside the areas of the figure on the right will be “not in proximity" to a new development.

and Gibler (2014). Rings are created around a new development, where inside the ring, a building would deemed as being “close to a new development." Buildings that fall outside the ring would be “not in proximity to the development”. The radius of these rings is variable that could be adjusted while determining which buildings are in the treatment and control groups. When experimenting with different radii, a smaller ring of 250m was used, as the density of new developments, especially in Brooklyn, was quite large. An example from a new development on the Upper West Side can be seen below

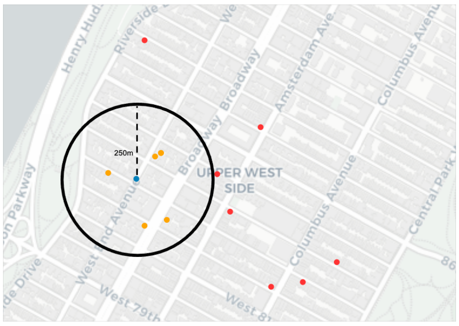

Figure 2: The graph above depicts a new development (in blue) from the HPD dataset in 2018 on the Upper West Side. The ring is drawn around the new development at a radius of 250m. The orange points represent building addresses that were randomly sampled inside the ring and thus are “in proximity" to the new development. The red points are outside of the ring and thus would not be in proximity to the new development.

Now that the definition of proximity and the data source of new developments have been explored, it is important to state the data source of the addresses of buildings used for this study. There are general datasets that contain all addresses in New York City; however, these datasets include office buidlings, retail and single family housing, which would need to be excluded from this study. The Housing Preservation Department provides a useful dataset, namely the Multiple Dwelling Registrations [2] (MDR) dataset, which contains all rental buildings in the city, as they need to register with the HPD in order to legally operate.

With the MDR dataset, the procedure to create the treatment and control groups in this study is as follows:

1. Find all areas considered for the treatment group and control group.

2. With these areas, find all addresses in each group and assign addresses as “in proximity" or “not in proximity".

3. Create equal samples of 1,000 buildings from each the treatment and control groups.

For the purposes of this study, as seen in step 3, only a sample of all buildings were used. The MDR dataset contains 179,030 unique buildings, which due to resource constraints, it would not be possible to web-scrape each building. Hence, a more manageable dataset was generated through sampling 1,000 buildings for the treatment and 1,000 for the

[2] Multiple Dwelling Registrations Dataset; Department of Housing Preservation and Development (HPD)

control group. With these 2,000 candidate buildings the next step is to attempt to web-scrape each of them to try to find attributes of the apartments.

5.2 Webscraping

Apartment details can be difficult to find, as there are no comprehensive rental datasets. For this reason, apartment attributes were obtained by web-scraping the apartment rental site StreetEasy. The unique aspect of the specific rental site is that it shows the full price history of an apartment, which usually is not public. This in turns allows for prices to be obtained for each apartment before and after the chosen completion year of 2018. Outlined in this section is the general process that was used to web-scrape apartments for the final dataset.

To begin with, there are three main components of the StreetEasy site that were used to gather attributes on apartments: the building search tool, building pages and apartment pages. The search function, which allows for queries directly from the website's search URL, allows a user to search for buildings based off of addresses. If there is an exact address match, the search function redirects to the building with the corresponding address. Hence, each building's address in the 2,000 sample dataset was searched for using this function. Samples that were not found, unfortunately, had to be discarded. The reasons why certain buildings did not appear on StreetEasy, could be due to address differences between the MDR dataset and the address listed on StreetEasy. Additionally, some rental buildings may use other platforms and thus would not be on StreetEasy.

Once a building page has been obtained, the page is scraped in its entirety. At the bottom of each building page is a list of apartments under "Units". Below is an example of the HTML table that was scraped:

Units

ACTIVE LISTINGS | PAST SALES | PAST RENTALS | ALL UNITS | Filter Units

Date	Unit	Rent	Beds	Baths	ft²	Floorplan
04/06/2021	#3G	$2,100	1 bed	1 bath	0 ft²	
03/08/2021	#1AA	$1,635	studio	1 bath	0 ft²	
02/20/2021	#4D	$2,100	2 beds	1 bath	0 ft²	FLOOR PLAN
02/19/2021	#4F	$2,100	2 beds	1 bath	800 ft²	
01/31/2021	#E3	$1,950	2 beds	1 bath	0 ft²	

Figure 3: As seen above, the table includes: the unit that has been rented, the price, and number of beds and baths.

The Units table in Figure 3 is important as it contains links to each apartment that has been rented. This study only used "Past Rentals", as apartment candidates. Active listings were not used, as a rental on the market may have price fluctuations that may not be representative of the price individuals are willing to pay for a specific apartment.

With the links to each apartment, the script then web-scrapes the apartment pages. The apartment page contains useful information such as the amenities and the price history of the apartment. The amenities appear under an "Amenity" section, which displays each amenity as plain text in HTML. Hence, the web-scraper gathers these amenities and stores them for a specific apartment.

The price history is an integral component of the data gather process for this study. Through the table, apartment prices before and after 2018 can be obtained. The caveat with the history is that each entry does not correspond to a rented apartment, as there may be de-listings or price changes. For the purposes of the study, these prices have been considered as equal; however, further research could improve upon this study by considering which types of listings should be used for the price model. Landlords can relist and delist apartments at various prices to test the markets, which could potentially add noise to the price change that this study is attempting to estimate.

Price History

12/11/2020	Listed by Douglas Elliman	$3,875
08/01/2017	Douglas Elliman Listing rented	$4,000
06/13/2017	Previously Listed by Douglas Elliman	$4,000
04/01/2016	Douglas Elliman Listing rented Last priced a...	$3,800
11/16/2015	Previously Listed by Douglas Elliman	$4,300
01/10/2011	Douglas Elliman Listing rented	$2,995
12/21/2010	Previously Listed by Douglas Elliman	$2,995

Figure 4: The table shows an exmple price history of an apartment. It provides the date of the listing, a description of why the price changed and finally the rental price.

Through the amenities and price history, the web-scraper is able to gather the data that may be conducive to changes in prices. It should be noted that this process removed the majority of buildings. Many buildings did not appear on search and other buildings that were found had no past rentals. Due to resource limitations, it was not feasible to run the scraper on another batch of buildings. In total there are $n = 123$ buildings which have apartments that satisfy the criteria of having prices before 2018 and specifically on 2019, as a post-completion year. Of these 123 buildings in total, there are 262 apartments for this study.

5.3 Data Analysis

The rental price of an apartment is influenced by multiple apartment attributes including location, size, and amenities. In our model, we want to control for other factors that could cause higher rent prices outside of the proximity to a new building development. Since the difference in differences estimator is attempting to estimate the causal effect of proximity on price, any confounding variables in price that are not controlled for in the regression may bias the estimator. However, it should be noted that some variables are quite sparse and hence we must justify their addition to the model, as the number of apartments to fit is small and the addition of sparse variables may add a large variance to our estimator.

Figure 5 displays the distribution of apartment rent prices in our data by borough and proximity to a new development. Our data set is bias towards Manhattan and Brooklyn, with very few apartments located in the Bronx and Queens. In addition, our sample contained no apartment units in Queens that were located close to a new development. We would expect apartments that are in proximity to new developments to have a higher median and range than those that are not in proximity. Figure 5 shows little difference in the price median between the two groups. This figure does not show any other factors that may influence the rental price, such as the number of bedrooms and building amenities, so we performed additional analysis for variables to use as control. Since each apartment in assigned a borough and location is a strong confounding factor in price, borough information should be controlled for in our model.

We then hypothesized that apartment amenities, such as a gym and in-unit laundry, would place an upward trend on rental price. Our data included a field that listed out the amenities available for each apartment. There were two issues we had to solve for with this field. The first is that there were many apartments that had no amenities listed. The second was that there were many amenity variables and few instances of each amenity. Therefore, we decided to test the hypothesis that apartments with at least one amenity have higher rental prices than ones with none listed. If there is a significant difference in price, then it is a notable confounding factor of price. Thus we can justify the additional noise that may be introduced to our estimator by controlling for such a granular variable.

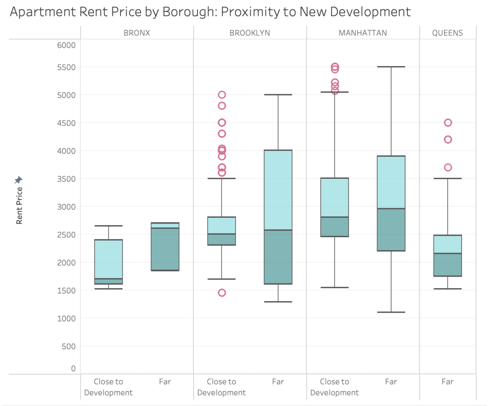

Figure 5: Rent Price by Borough

To test this hypothesis, we ran a t-test on apartment prices. In total there were 138 apartments with no amenities and 124 apartments with at least one amenity. A pair sampled t-test was conducted to detemine the effect of the existence of at least one apartment amenity on rent price. There was a significant difference between apartments with at least one amenity (M=3646, SD=2782) and apartments that had no amenities listed (M=2656, SD=720). Despite low occurrence of each amenity, this test showed that a binary flag for the presence of an amenity does have a significant effect on rental price, and therefore was used in our model to control for this factor.

Figure 6: Rent by Number of Bedrooms

Lastly, another factor we considered as a con-

Factors	Coefficient	p-value	0.025 CI	0.975 CI
Number of Bedrooms	573.27	0.00	490.06	656.48
Number of Baths	-77.62	0.48	-292.35	137.114
Amenities	600.77	0.00	404.25	797.30
Time (T_i)	-194.28	0.12	-438.73	50.17
In Proximity ($PROX$)	-606.34	0.00	-841.83	-370.85
Diff. in Diff. ($PROXT_i$)	284.86	0.07	-26.01	595.74
Manhattan	2481.97	0.00	2192.64	2771.30
Bronx	1306.60	0.00	659.07	1954.14
Brooklyn	1902.79	0.00	1574.44	2231.14
Queens	1340.11	0.00	948.37	1731.85

Table 1: The table above shows the results from the hedonic model fitted to our dataset.

founding factor in price would be the number of bedrooms and bathrooms of an apartment. Both variables showed an upwards trend in rental price. Specifically, Figure 6 shows an upward trend in rent price as the number of bedrooms increased from a studio to 5 bedrooms. Since, both variables were not sparse, as each apartment had an assigned number of bedrooms and bathrooms, we included them in the model as a confounding factor in price.

6 Results

The dataset consists of the following columns corresponding to the factors which were considered for determining the price of an apartment:number of bedrooms, number of baths, amenities flag (binary valued: 1 if amenities were listed, 0 otherwise), pre_flag (binary-valued: 1 if the apartment was completed before 2018, else 0), in-proximity (binary-valued: stores the value 1 if the apartment is close to a new development and 0 otherwise), and four one-hot encoded columns for each of the boroughs. The implementation of the difference in differences estimator was carried out by constructing a column which stores the product of the binary valued columns of 'in-proximity' and 'pre_flag'. From equation (1), the *PROX* variable corresponds to the 'in-proximity' column and the 'pre_flag' column represents the T_{it} variable.

In order to observe the effects of the various factors on the price of an apartment a linear regression model was constructed. The results of the constructed model are presented below.

6.1 Model Results

The coefficient of $PROXT_i$ (θ from equation (1)) is relatively low compared to the other factors. It also has a large variation in the 97.5 confidence interval and a p-value of 0.072, implying proximity to new developments does not play a significant role in determining prices. For the same reasons, number of baths and completion time of an apartment are not statistically significant. The coefficients corresponding to the number of bedrooms and presence/absence of amenities have a positive correlation on the prices of apartments which is expected. The variable indicating the proximity to new developments has a negative correlation on prices which is in contrast to the coefficient of $PROXT_i$ variable thus resulting in an inconclusive correlation. The columns corresponding to the Manhattan and the Brooklyn columns have high coefficients because a large fraction of the apartments in the dataset are from these boroughs.

The model achieves an R^2 value of 0.413 on the dataset. This can be attributed to the fact that the model is unable to capture the high variance values displayed by several of the variables which are considered for determining the price.

7 Discussion

While, the difference in differences estimator has too much variance to accurately deduce whether there is a statistically significant relationship between new developments and rental prices, there is much room for future research to improve upon the methods created in this study and gather insights to the causal relationship at hand.

The regression model shows that a new development after completion has a slightly positive effect on surrounding apartment buildings at a confidence level of 90%. Furthermore, it appears that the 90% confidence level should not be used in this case, since the model has a fairly low R^2 value. Although this study cannot conclude there is a relationship between new development and surrounding rental prices, it appears to be a possibility with further research necessary. The model is not able to explain the variance in price well, as there was not enough data in the study and there could be other important factors for rental prices, which could not be obtained for this study.

As mentioned in the data section, the data sample included 123 buildings and 262 apartments. To reduce the variance of our difference in differences estimator, future work will need to include a larger sample of apartments, which would include data from sources other than StreetEasy. To gather a larger dataset, more robust approaches to web-scraping the rentals sites should be used. Towards the end of the scraping session of the initial sample of 2,000 apartments, the error rate in web requests grew exponentially until it was not feasible to gather anymore apartments. Changing IP addresses more often and requesting apartment URLs in a more randomized fashion may help alleviate some of these issues that arose during the data gathering portion of this study. Additionally, more complex heuristics can be created in terms of matching buildings to search results. This study only used exact matches between addresses and StreetEasy listings, which could potentially have constricted the number of buildings found. Using natural language processing tools, an address matching system could be developed to select the correct result from an address search, if the addresses are a match but are written differently.

In addition to the small sample of apartments, the data in this study are biased towards the Manhattan and Brooklyn rental markets. Of the 262 apartments that met the criteria of price listings in the correct year, 132 apartments are located in Brooklyn and 108 are located in Manhattan, leaving a total of 22 apartments located in Queens and the Bronx. The discrepancy of the number of apartments in different boroughs could be attributed to multiple factors. Since the distribution of boroughs from the web-scraped data does not match that of the 2,000 candidate building dataset or the distribution of new developments found in Figure 1, it would seem that there was a higher error rate in finding apartments on StreetEasy in boroughs besides Brooklyn or Manhattan. The systematic difficulty in finding apartments in Queens, the Bronx and Staten Island on Streeteasy may be due to a difficulty in matching addresses from the MDR dataset. In addition, since Queens, the Bronx and Staten Island tend to have lower density zoning, it may be the case that the buildings from the MDR dataset may be detached homes and thus separate apartments are intended for family member use and hence would not be listed on StreetEasy.

Since price history from StreetEasy as an integral part of the data in this study, future work will include more analysis into this variable on StreetEasy. The historical entries on the site are not guaranteed to represent the current price residents are paying in rent. As previously mentioned, there be de-listings or re-listings where landlords may be testing the rental markets in the surrounding area. A more accurate representation of what tenants were paying in rent historically could also improve the results of the study and show a more accurate estimate of the difference in differences.

In terms of the parameters when choosing candidate buildings for treatment and control groups, more research needs to be done on the best size of ring for neighborhoods in New York City. This study used a ring of 250 meters, which is smaller than those used in (Ooi and Le, 2013) and (Zahirovich-Herbert and Gibler, 2014). This choice was made, as the prior work has been done in less densely populated neighborhoods compared to areas with multiple dwelling buildings in New York City. Density could play a role in dampening price changes due to new construction, as each new addition to a neighborhood becomes less noticeable and thus amenity effect is less apparent. Adjusting the ring size could provide better results. Specifically, future work should try smaller ring sizes, which could yield treatment groups that are closer to the new developments and thus may have a more statistically significant effect on the treatment group's rental prices.

This study followed a similar method that (Ooi and Le, 2013) used in their research on housing sale prices in Singapore. In addition to a population level treatment and control group, their model accounted for specific neighborhood price variance, introducing another variable to indicate various areas of Singapore. Unlike in prior work, the difference in differences estimator, in this study, had a negative effect on rental price. Even though it is not a statistically significant result, it is important to understand how the treatment and control groups may have played a role in the negative coefficient.

One hypothesis is that there may be a bias of higher rental prices in the control group. Since, Manhattan had a lower error rate when web-scraping StreetEasy and since a large portion of Lower Manhattan is not in proximity to new developments as seen in Figure 1, Lower Manhattan is over-represented in the apartments scraped. This can be seen in Figure 5, where buildings in Manhattan far from new developments seem to have a higher range of rental prices. The lack of development in more expensive neighborhoods, specifically in Lower Manhattan, can be attributed to the historic nature of the area and the difficulty of getting construction permits from the Landmark Preservation Commission (LPC)[3]. The LPC's main role is to preserve historic architecture, which may be lost with the construction of new developments. To control for this neighborhood level bias and reduce variance in the difference in differences estimator, further work should attempt to include more granular data on the location of an apartment. Specifically, a neighborhood variable in addition to the borough variable should help reduce the variance in the data. With the addition of a granular variable such as neighborhoods, it would be necessary to have apartments in the control and treatment group in every neighborhood in order for the quasi-experiment to be accurate.

Lastly, since finding a complete list of market-rate new developments constructed in New York City was difficult, this study used the HPD dataset which includes mixed affordable and market-rate housing. It is not clear whether the effect of mixed-income development on surrounding rental prices would be different from the effect of market-rate housing. However, future work should attempt to gather buildings that only have market-rate housing in order to make stronger conclusions of the price effect of market-rate housing.

8 Conclusion

Through this study a method of gathering relevant data and creating an initial experiment was outlined to estimate the effect that new developments have on surrounding rental prices. While the resulting effect was not statistically significant, the methods and analysis in this study can be used to further our understanding of the effect of new developments and in turn inform fairer housing policy for residents of all incomes.

[3]Landmark Preservation Commission; The about page contains information about their purpose and information.

References

Katrin B. Anacker. 2019. Introduction: housing affordability and affordable housing. *International Journal of Housing Policy*, 19(1):1–16.

Brian J. Asquith, Evan Mast, and Davin Reed. 2019. Supply shock versus demand shock: The local effect of new housing in low-income areas. *Upjohn Institute*, 8(3).

Vicki Been, Ingrid Gould Ellen, and Katherine O'Regan. 2019. Supply skepticism: Housing supply and affordability. *Housing Policy Debate*, 29(1):25–40.

Jack Buckley and Yi Shang. 2002. Estimating policy and program effects with observational data: The 'differences in differences' estimator. *Practical Assessment, Research, and Evaluation*, 8(2).

Ingrid Ellen, Michael H. Schill, Amy Ellen Schwartz, and Scott Susin. 2001. Building homes, reviving neighborhoods: Spillovers from subsidized construction of owner-occupied housing in new york city. *Journal of Housing Research*, 12(2):185–216.

Alicia Glen. 2014. Housing new york: A five-borough, ten-year plan.

Xiaodi Li. 2019. Do new housing units in your backyard raise your rents. Technical report.

Joseph T.L. Ooi and Thao T.T. Le. 2013. The spillover effects of infill developments on local housing prices. *Regional Science and Urban Economics*, 43(6):850–861.

Velma Zahirovich-Herbert and Karen M. Gibler. 2014. The effect of new residential construction on housing prices. *Journal of Housing Economics*, 26:1–18.

9 Summary of Work

We all participated in weekly planning sessions and meetings to determine our plan for the project and how we were going to pull data, create the proximity variable, and clean the data for our final model. Sotiris created the web-scraping script and scraped the data from StreetEasy. Kerry performed the initial data analysis and cleaned the data for the final model. Nripesh created the modeling notebook and ran different model tests. We each wrote the parts of the paper that corresponded with the analysis performed, and all reviewed the entire paper for fluency and spelling/grammar.