Taxi Activity as a Predictor of Residential Rent in New York City

by

Philip Caporaso

B.S., Civil Engineering, 2013

Purdue University

Submitted to the Program in Real Estate Development in Conjunction with the Center for Real Estate in Partial Fulfillment of the Requirements for the Degree of Master of Science in Real Estate Development

at the

Massachusetts Institute of Technology

September, 2019

Signature of Author____
Center for Real Estate
July 1, 2019

Certified by____
Alex van de Minne
Research Scientist
Center for Real Estate
Thesis Supervisor

Accepted by____
Professor Dennis Frenchman
Class of 1922 Professor of Urban Design and Planning,
Director, Center for Real Estate
School of Architecture and Planning

This page intentionally left blank

Taxi Activity as a Predictor of Residential Rent in New York City

by

Philip Caporaso

Submitted to the Program in Real Estate Development in Conjunction with the Center for Real Estate on July 1, 2019 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Real Estate Development

ABSTRACT

Real estate developers and investors have a vested interest in discovering new techniques for estimating the direction and magnitude of changes in residential rent within a neighborhood. This study hypothesizes, and finds evidence, that taxi activity is a proxy for changing income and neighborhood quality as well as an indicator of gentrification.

Novel research is performed to determine if taxi activity is a significant predictor of rents in New York City at the neighborhood level. Nine OLS regression models are created using data about 1,466,234,991 taxi pickups and drop-offs, median rent, and median income across 188 neighborhoods in New York City in the years of 2010-2015. In all nine models, taxi activity is found to be a statistically significant predictor of rent at 99% confidence.

This study finds that a 1 standard deviation positive shock in taxi drop-offs will result in a 0.009% - 0.155% higher rent the next year on average.

Thesis Supervisor: Alex van de Minne
Title: Research Scientist

Table of Contents

1. INTRODUCTION **5**

2. LITERATURE REVIEW **6**

2.1 Residential Rents and Prices 6
2.2 Gentrification 8
2.3 Taxis 9

3. METHODOLOGY **10**

4. DATA **13**

4.1 Spatial Data 13
4.2 Taxi Trip Data 13
4.3 Median Income Data 15
4.4 Median Rent Data 16

5. RESULTS **18**

5.1 Parametrization of the Model 18
5.2 One-Year Lagged Linear Regressions 19
5.3 One-Year Lagged Rent 19
5.3 One-Year Lagged Trips 20
5.4 One-Year Lagged Income 20
5.5 Dummy Year and Borough Variables 21
5.6 Model Fit 21
5.7 Discussion 22

6. CONCLUSION **26**

7. REFERENCES **28**

1. **Introduction**

The research contained within this study was performed in order to determine if taxi activity has predictive power for determining changes in residential rents at the neighborhood level in New York City. Ordinary Least Squares regression models are created from panel data to determine the significance of taxi activity in the context of residential rents and incomes.

Prior research on residential rents indicates that median income and neighborhood quality are significant determinants of rent. Recent research describes gentrification as, among other things, an influx of younger and wealthier residents, leading to higher rents. Taxi users in New York City are on average younger and wealthier than the general population of New York City.

In light of the fact that taxi users are more likely to match the profile of a person in a gentrifying neighborhood than a randomly selected person of the general population, and the fact that gentrification is associated with increasing income and neighborhood quality, it is hypothesized that an increase of taxi trips in a neighborhood are a proxy for coming gentrification. If a neighborhood experiences an increase in taxi trips over a period, then one year later, quality and income of that neighborhood increase, and therefore rents of that neighborhood are expected to increase. In the inverse case where taxi activity declines, rents are expected to decline.

Panel data for this study is obtained from New York City's Taxi and Limousine Commission and the American Community Survey, as published by the New York City Department of City Planning. The taxi and rent data suffer from issues that may cause underestimation of the effects of taxi activity – the taxi data does not include ride-hailing services such as Uber and Lyft and the median rent data does not control for structural characteristics of housing and includes rents for housing under some forms of rent control which do not respond to market forces.

Three regressions are performed three times each to test three definitions of taxi activity: the annual number or taxi pick-ups, the annual number of taxi drop-offs, or the annual number of pick-ups and drop-off in a neighborhood. The first set of models includes dependent variables of time lagged rent, time lagged taxi activity, time lagged median income, dummy variables for the borough containing each neighborhood, and dummy variables for each year examined. Next, time lagged median income and the dummy variables are excluded and the regression is performed again. Last, time lagged median rent is excluded and the regression is performed solely on taxi activity.

All of the regressions performed indicated that time lagged taxi activity is significantly and positively correlated with median rent at 99% confidence. Drop-offs were found to provide the best model fit, when rents were regressed solely on taxi activity. The effect of a 1 standard deviation positive shock in taxi pick-ups drop-offs will result in a 0.009% - 0.150% higher rent the next year on average.

This research is concluded with a discussion of the potential issues of the data used and suggests further areas of research.

2. Literature Review

Through the time of writing of this paper, no research has been published that attempts to relate residential real estate prices or rents to taxi activity. However, residential real estate prices and rents have been extensively researched. Similarly, taxi usage has been extensively studied in recent times as data has become available over the last decade.

2.1 Residential Rents and Prices

Rosen (1984) established the hedonic pricing model. He found that “A class of differentiated products is completely described by a vector of objectively measured characteristics”. While the model was a general economic model, it has since been applied extensively to real estate. Examples of hedonic characteristics with respect to residential real estate are features such as size, distance from a point of economic interest, number of bedrooms, number of bathrooms, amenities, and such other characteristics.

Sirmans and Benjamin (1991) summarize the prior work conducted on creating hedonic pricing models for real estate. The literature summarized found that a multitude of factors have statistically significant effects on rent. Some notable variables with positive effects include population density, laundry services, size, number of bedrooms, number of bathrooms, professional property management, and length of tenancy. Other variables, such as building age, vacancy rate, traffic congestion, distance from Metropolitan Statistical Area, and adjacency to a major thoroughfare had negative effects.

Median Income was found to be most significant determinate of rent by Gilderbloom and Appelbaum (1987). The primary purpose of their research was to determine if "urban rental markets are competitive and that rent levels are determined solely by the supply and demand for housing". Their methodology was to create standardized and unstandardized regression analyses of median rent in 140 urban markets using 12 independent variables. Of these variables, median income and median house cost were found to be significant at 99% confidence in both regressions. Both determinates were found to have a strongly positive correlation. Restated, areas with higher incomes and higher house costs pay higher rents.

Of the variables they studied, the correlation of median rent with median income was found to be the strongest in both regressions. The relationship "implies that, in many housing markets, landlords are in fact able to charge what the market will bear--even when differences in demand and quality are taken into account". Further, "[t]he purely frictionless supply and demand model cannot adequately why rents rise. The social factors [Gilderbloom and Appelbaum] have examined in this paper have significance on rent levels – influences that economists have not taken into account". This finding is more pronounced in areas where ownership is highly professionalized, that is where rental stock is owned by sophisticated investors.

Neighborhood effects on rent are considered by Marks (1984) while studying the effects of rent control on the hedonic pricing model in the City of Vancouver. He found that "'better' neighborhoods have higher rents". In the study a "better neighborhood" is one on the westside of Vancouver and is represented in the regression as a categorical variable for east or west. Unsurprisingly, neighborhood desirability has no statistically significant effect on controlled rents. However, uncontrolled units have a strong positive correlation with a westside location at 99% confidence. The research indicates that neighborhood desirability has a strong and significant effect on residential market rents.

Neighborhood quality is difficult to objectively observe. Dubin (1992) addressed this issue by delineating rents into structural, neighborhood quality, and accessibility components. The structural term was the physical attributes that are common in most hedonic regressions, such as those described above. The neighborhood quality and accessibility components were then contained in the error term and their autocorrelation is modeled. Dubin emphasized the issue of neighborhood boundaries being non-definitive and that a fixed boundary is often an inaccurate method of defining a neighborhood. Using kriging, sizable neighborhood effects were demonstrated in Baltimore. Spatial autocorrelation reveals that certain areas have location premiums, while others have penalties that are not consistent with the monocentric city model.

2.2 Gentrification

The term gentrification was first coined by Ruth Glass in 1964 to refer to a phenomenon in London where working class neighborhoods were being transformed by an influx of lower and upper middle class residents. This caused the housing prices to increase to a level that could not be supported by its existing residents.

There are 4 stages of gentrification (Clay, 1979).

1. "[A] small group of risk-oblivious people move in and renovate properties for their own use."
2. "[A] few more of the same type of people move in and fix up houses for their own use. Subtle promotional activities are begun, often by a few perceptive realtors."
3. "[M]ajor media or official interest is directed to the neighborhood… The arrivals in this third stage include increasing numbers of people who see the housing as an investment in addition to being a place to live."

4. "[A] larger number of properties are gentrified, and the middle-class continues to come. What is significant about the new residents is that more are from the business and managerial middle class than from the professional middle class."

Today, the term gentrification is not precisely defined but is used in most contexts to refer to the change of a neighborhood from primarily consisting of low-income residents to primarily consisting of new higher income residents.

Of importance, gentrification occurs in particular neighborhoods and not across an entire city. When a positive shock to demand occurs, gentrifying neighborhoods experience a greater price increase than the rest of the city (Guerrieri et al 2013). The variation of house price growth between neighborhoods within a city is large. Neighborhoods with lower prices prior to the demand shock tend to appreciate greater, with higher variation, than neighborhoods with higher initial house prices. Poor neighborhoods close to wealthy neighborhoods tend to appreciate more than poor neighborhoods that are distant from wealthy neighborhoods.

Prediction of future gentrification is possible. Behrens et al (2019) identified "pioneer businesses" that are over represented in poor but soon-to-gentrify neighborhoods. These businesses are from "mostly cultural, recreational, and creative industries". Their over representation in poor neighborhoods "foster gentrification through the types of workers they hire, their signal as to the future prospects of a neighborhood, and their effect on the subsequent arrival of consumption amenities." These pioneer businesses do not provide goods and services that can be consumed by the residents of the neighborhoods where they operate, suggesting an inflow on non-residents to sustain the businesses. Pioneer businesses also precede the arrival of consumption amenities such as bars and restaurants which may trigger the arrival of new affluent residents.

2.3 Taxis

In 2014, New York City's Taxi and Limousine Commission released an updated Taxicab Fact Book. The most recent fact book prior to 2014 was released in 2006. Since 2014, an updated fact book has been released every 2 years. In 2014, a year contained in the study's data sample, several facts are revealed. First, 70% of all taxi passengers are aged 35 or younger. Half of those passengers were age 20 or younger. These age groups are over represented as the general population of this group is 51%. Next, 42% of taxi riders have an income of $100,000 or greater. Citywide, this income group only represents 24% of the population, however in Manhattan, where the majority of taxi trips occur, 37% of the population has an income of $100,000 or greater. Generally speaking, taxi users in New York city are younger and have a higher annual income than the general population.

3. Methodology

Median rent is modeled by regression analysis with a one year time lag of the independent variables. Lagged independent variables are used because median rent cannot move contemporaneously with exogenous changes. Construction of new units is not instantaneous, nor can landlords instantaneously increase rents when tenants are subject to a residential lease of a term which varies from renter to renter.

Because no prior research has been published, the best definition of taxi activity is not yet known. In order to test for the most significant definition, each form of model will be performed three times using one of three different definitions. The three definitions studied are the number of taxi pickups, drop-offs, or pickups and drop-offs.

Nine models will be created from three unique equations by utilizing panel data from 188 neighborhoods in an Ordinary Least Squares regression model, whose general form is:

$$Y = \propto_0 + \propto_1 X_1 + \propto_2 X_2 + \cdots + \propto_n X_n + \varepsilon \qquad (3\text{-}1)$$

Where Y is the dependent variable, $\propto_0$ is the intercept term, X_1 through X_n are the independent variables , $\propto_0$ through $\propto_n$ are the regression coefficients, and ε is an error term.

The first set of models will use three equations of the form:

$$\widetilde{\mathrm{RENT}}_{\iota,t} = \propto_0 + \propto_1 \widetilde{\mathrm{RENT}}_{\iota,t-1} + \propto_2 \widetilde{\mathrm{TRIPS}}_{\iota,t-1} + \propto_3 \widetilde{\mathrm{INC}}_{\iota,t-1} + \propto_4 2012 + \propto_5 2013$$
$$+\propto_6 2014 + \propto_7 2015 + \propto_8 Brooklyn + \propto_9 Manhattan + \propto_{10} Queens$$
$$+ \propto_{11} Staten\ Island + \varepsilon$$

(3-2), (3-3), (3-4)

where:

$\widetilde{\mathrm{RENT}}_{\iota,t}$	is the log transformed median rent in neighborhood i at time t
$\widetilde{\mathrm{RENT}}_{\iota,t-1}$	is the log transformed median rent in neighborhood i at time t – 1
$\widetilde{\mathrm{TRIPS}}_{\iota,t-1}$	is the log transformed amount trips in neighborhood i at time t – 1. Trips are the number of pickups, drop-offs, or pickups and drop-offs in each equation, respectively
$\widetilde{\mathrm{INC}}_{\iota,t-1}$	is the log transformed median income in neighborhood i at time t – 1
$2012\ etc$	are dummy variables for year t
$Brooklyn\ etc$	are dummy variables for the borough in which neighborhood i is located

RENT, INC, and TRIPS are log transformed values for ease of understanding the regression results. Rent and income are of the magnitude of thousands, while trips vary significantly in magnitude from thousands to millions. The log transformation normalizes the data and should result in coefficients of similar magnitude.

The second set of models will exclude time lagged median income and the dummy variables and take the form of:

$$\widetilde{\mathrm{RENT}}_{\iota,t} = \propto_0 + \propto_1 \widetilde{\mathrm{RENT}}_{\iota,t-1} + \propto_2 \widetilde{\mathrm{TRIPS}}_{\iota,t-1} + \varepsilon \qquad (3\text{-}5), (3\text{-}6), (3\text{-}7)$$

where the variable definitions are the same as the first set of models.

The last set of models exclude time lagged median rent and take the form of:

$$\widetilde{\mathrm{RENT}}_{i,t} = \propto_0 + \propto_1 \widetilde{\mathrm{TRIPS}}_{i,t-1} + \varepsilon \qquad (3\text{-}8), (3\text{-}9), (3\text{-}10)$$

where the variable definitions are the same as the first two sets of models.

The RENT and INC variables have gaps in the data set where their values were not reported during certain years. These variables were assumed to change on a straight-line basis using:

$$X_y = X_t + (y - t)\frac{X_z - X_t}{z - t} \qquad (3\text{-}11)$$

where X_y is the unreported value in year y, X_t is the most recent prior reported value which is in year t, and X_z is the next reported value which is in year z.

RENT is selected as a time lagged dependent variable current as future rent is expected to be a change from the prior value. The coefficient on RENT is therefore expected to be positive.

TRIPS is the variable of interest in this study. If the hypothesis is correct, an increase in TRIPS will correspond with increasing neighborhood quality and a later increase in median income (gentrification). The inverse is expected to be true, where decreasing trips corresponds with decreasing neighborhood quality and income. Therefore the coefficient on trips change is expected to be positive.

INC is included as an explanatory variable, drawing from Gilderbloom and Appelbaum's (1987) assertion that residential rents are not efficient, but rather priced to what the market can bare. It is expected that the coefficient will be positive.

2012 etc and Brooklyn etc are included to account for fixed effects that occur across the entire sample in a year, or in a particular borough. The years are expected to have positive coefficients because rent has trended upward throughout the time period of data that was analyzed. The boroughs are expected to be either positive or negative. No prediction is made for the sign of the coefficients for boroughs. In all models, the year of 2010 is dropped because no lagged data was generated. 2011 is the omitted year variable and The Bronx is the omitted borough variable.

4. Data

4.1 Spatial Data

New York City's Department of City Planning (DCP) has established boundaries for Neighborhood Tabulation Areas (NTAs). NTAs were created to roughly establish neighborhoods containing a minimum of 15,000 residents each. While NTAs do not exactly correspond to historic neighborhood boundaries, they are the best available approximation of neighborhoods for which economic and housing data is available.

In total, DCP has defined 195 NTAs where there are: 51 NTAs in Brooklyn, 38 NTAs in The Bronx, 29 NTAs in Manhattan, 58 NTAs in Queens, and 19 NTAs in Staten Island.

4.2 Taxi Trip Data

Metering for NYC taxis is provided by Technology Service Providers (TSPs). Beginning in 2009, New York City's Taxi and Limousine Commission (TLC) began reporting certain information about yellow taxi trips, provided by the TSPs. Important to this research were the pickup and drop-off locations in the form of latitude and longitude coordinates and time of pick-up for every licensed taxi trip that were reported by the TSPs and published every 6 months by the TLC. Data is published as comma separated values file for each month, wherein each line of the dataset represents a single taxi trip.

In 2011, Uber, and other application based services such as Lyft, colloquially referred to as for-hire-vehicles (FHVs), began operating within New York City. Until 2015, the TLC did not have access to, or did not publish, any data relating to these types of taxis. Beginning in 2015, TLC began publishing this data, but only the trips' pick-up times and pick-up locations, in the form of a "taxi zone", were published. Drop-off location and drop-off time information did not begin to be published until the summer of 2017, and only for future trips. Because of the mismatch in the period and type of location data available for FHVs, FHV trips are not included in the dataset.

In August 2013, green taxis were introduced to NYC. The data reporting for green taxi trips began concurrently with the introduction of green taxis and exactly matched the type of data provided for yellow taxi trips.

In July 2016, the TLC changed the type of location reporting across all taxi types. Instead of providing a longitude and latitude coordinates, reporting was changed to "taxi zones", as is reported for FHV trips. Taxi Zones are artificial boundaries that roughly align with common designations of neighborhoods, similar, but not exactly the same as NTAs. Because economic and housing data is not available by taxi zones, taxi trips reported with pick-ups and drop-offs in taxi zones are not included in the data set.

Because of the mismatched and modified reporting of taxi trip information, yellow and green taxi trips in 2010, 2011, 2012, 2013, 2014, and 2015 are used for analysis as a consistent data set.

To prepare the taxi data for analysis, a count of pick-ups and drop-offs in each NTA much be calculated. First, all yellow and green taxi trip records from the years 2010 to 2015 are downloaded by month. Next each trip pick-up and drop-off location is evaluated to determine if it falls within an NTA. If the location falls within an NTA, the NTA code is recorded in the file; if the location is not within an NTA, a null value is recorded. For example, a trip that originates within Manhattan and terminates in New Jersey would record a pick-up NTA to that trip, and the drop-off NTA would be recorded as null.

Once each trip has been assigned with a pick-up NTA and/or a drop-off NTA, a count of the drop-offs and pickups that occurred in each NTA in each month is calculated. Next, the totals of pick-ups and drop-offs within each NTA, across yellow and green taxi trips, are summed into yearly totals. In total, this analysis located 764,684,876 drop-offs and 766,648,472 pickups within New York City, a total of 1,531,333,348 data points.

Next 7 NTAs with the suffixes of 98 and 99 are removed. These suffixes represent airports and all undevelopable land, such as parks and cemeteries but not vacant lots, respectively. Once these neighborhoods are removed, the data set consists of 738,016,749 drop-offs and 728,218,242 pick-ups for a total of 1,466,234,991 data points, a reduction of 65,098,357 data points.

The data set is then grouped by NTA and year for a final data set of 1,128 observations – six years of observations for each of the 188 NTAs. Examining the total number of pickups and drop-offs in an NTA in a year yields a mean of 1,301,007 trips, a standard deviation of 4,438,637 trips, a median of 37,531 trips, and a range of 44,467,390 trips. Figure 4-1 shows that the range and variance of the data can be explained by the larger magnitude of trips in Manhattan.

Log transforming the trip data, for pick-ups and drop-offs, normalizes it and yields a mean of 4.8, a standard deviation of 1.04, a median of 4.57 trips, and a range of 4.94. Plotting the histogram shown in Figure 4-2 reveals that the log transformed trips are roughly normally distributed in each borough, but not normally distributed as a single data set.

Lastly, viewing the percent change in trips by borough and year, as shown in Figure 4-3, illuminates that in 2014 the boroughs other than Manhattan experienced a large increase in the number trips, likely due to the introduction of green taxis in the second half of 2013.

4.3 Median Income Data

The New York City Department of City Planning publishes annual median income data delineated by NTAs. The original data is obtained from the American Community Survey, which is conducted by the United States Census Bureau. No data at the NTA level is provided for the year of 2011. Median income has been assumed to increase on a straight-line basis between 2010 and 2012 according to equation (3-11) to allow for analysis including income data in that year.

4.4 Median Rent Data

Like median income data, annual median rent data at the NTA level is obtained from American Community Survey data published by the New York City Department of City Planning. Median rent data for the years 2011 and 2012 is not available at the NTA level. As is assumed in the median income data, it is assumed that median rents increased on a straight-line basis for this period and calculated according to equation (3-11). Additionally, until 2015, the maximum median rent recorded by the American Community Survey was $2,000 per month. A median rent equal to $2,000 occurred in one neighborhood in 2010, 2011, and 2012, eight neighborhoods in 2013, and nine neighborhoods in 2014. It is inferred that a median rent equal to 2,000 underestimates the actual median rent in these neighborhoods.

Upon inspection, median gross rents appear lower than suggested by the researcher's anecdotal experience. The underestimation can be explained by two factors. First, the ACS median rent estimate uses the midpoint of reported gross rent for calculation and does not control for housing quality or unit size (Renwick, 2011). Accordingly, the median rent reported may refer to non-comparable units across neighborhoods. Secondly, in 2012 only 39.0% of rental housing units in New York City were market-rate rentals (NYU Furman Center, 2013). The remaining 61.0% of rentals were either rent stabilized, rent controlled, public housing, or subject to other forms of public subsidy which deflates the median rent from the market rent, and the response of median rent to changing market conditions.

While the actual figure of median rent may be underestimated, the inclusion of market rate rents in the data should cause the median rent to generally trend with market rent movements.

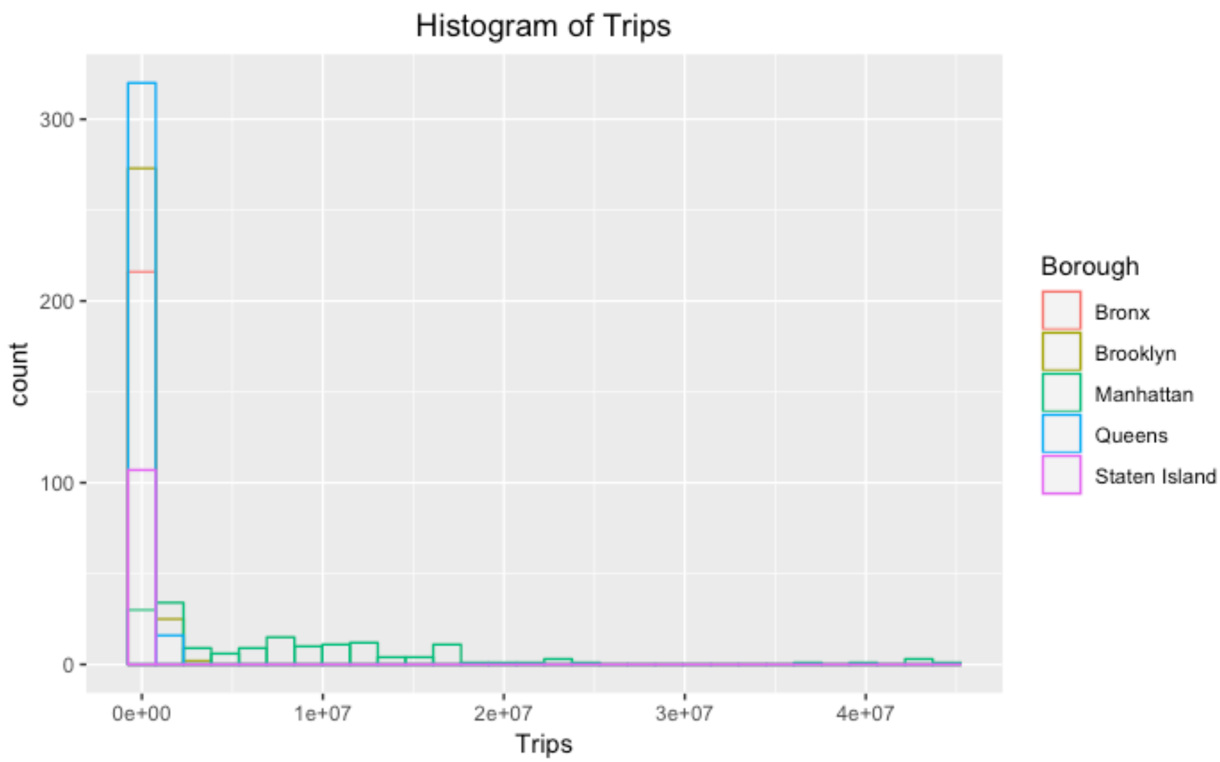

Figure 4-1

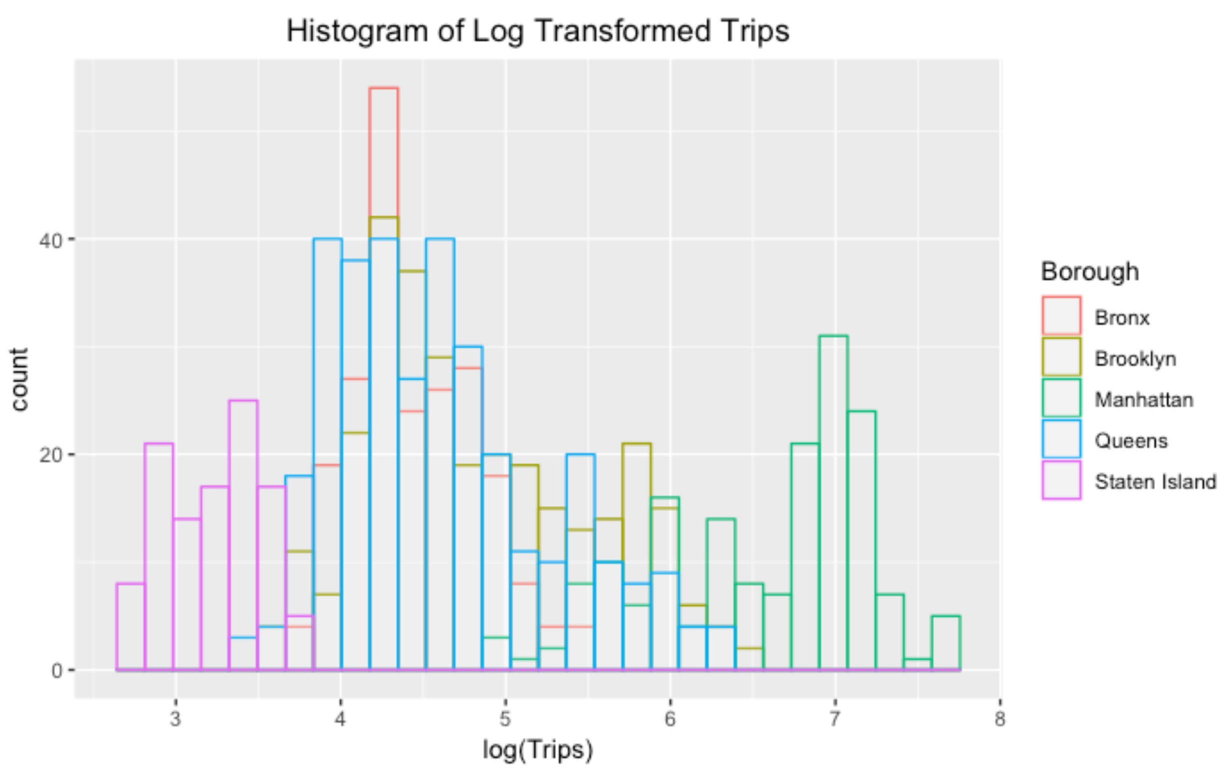

Figure 4-2

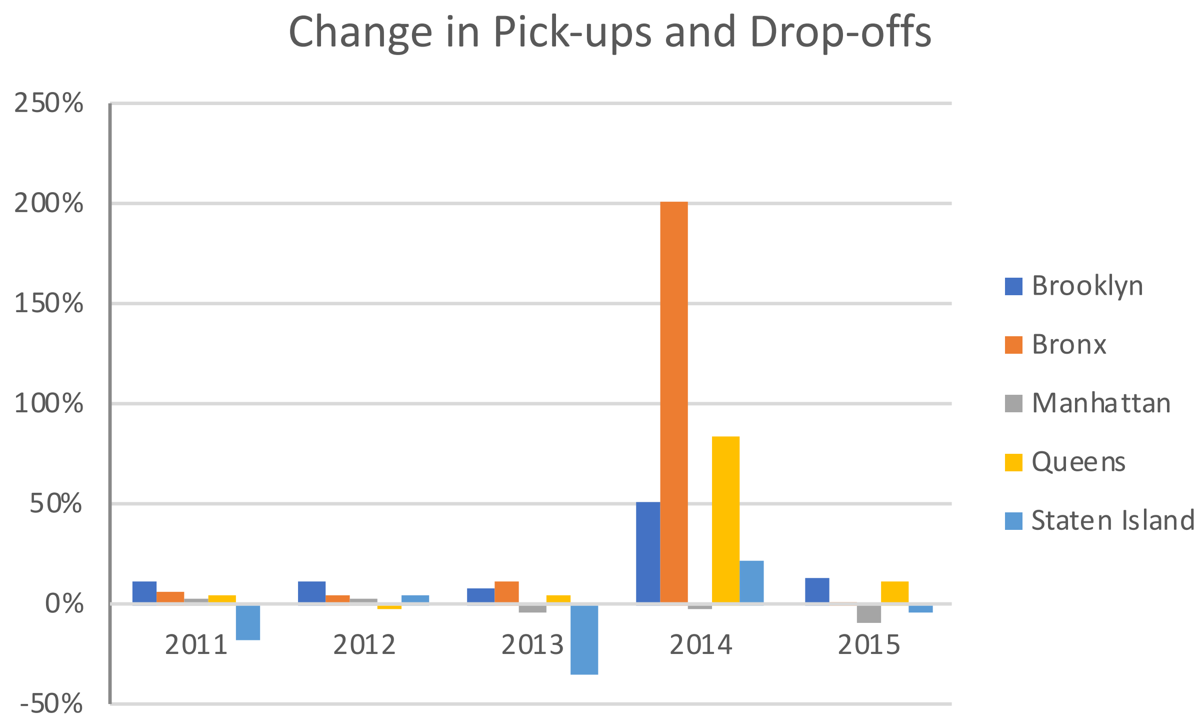

Figure 4-3

5. Results

5.1 Parametrization of the Model

As discussed in Section 4 (Data), and shown in Figures 4-1 through 4-3, the trips data is normally distributed only when log transformed and grouped by borough. Further, there is a significant change in the number of trips in the boroughs other than Manhattan in 2014. To account for the differing magnitudes of trips by borough and year, the anticipated fixed effects of the same, and standardization of data across all variables, the continuous independent variables are demeaned by:

$$\widetilde{x_{i,t}} = x_{i,t} - \bar{x}_{i,t} \quad (5\text{-}1)$$

where $\widetilde{x_{i,t}}$ is the demeaned variable for neighborhood i at time t, $x_{i,t}$ is the observed variable in neighborhood i at time t, and $\bar{x}_{i,t}$ is the mean of the variable in the borough containing neighborhood i at time t.

5.2 One-Year Lagged Linear Regressions

Table 5-1 shows the results of the linear regression to approximate the log transformed rent. The regression output indicates that median rent can be almost entirely predicted by the median rent of the prior year, the median income of the prior year, and the taxi activity of the prior year.

Table 5-2 shows that removing lagged median income and dummy variables from the regression does not have a significant effect on the model fit

Table 5-3 shows that time lagged rent was the best predictor of current rent, but also that it absorbed some of the predictive power of time lagged trips.

In all nine models, taxi activity is found to be a statistically significant predictor at 99% confidence.

5.3 One-Year Lagged Rent

$RENT_{t-1}$has a strongly positive correlation as expected. This indicates that RENT at time t is dependent on RENT at time t-1. This result matches general intuition that rent typically varies modestly from its prior value. In equation (3-2), the coefficient is slightly larger than it is in equations (3-3) and (3-4). This is likely because the TRIPS dependent variable used is pick-ups and has a lower coefficient, the significance of which is discussed below. Across all three equations, the coefficient on $RENT_{t-1}$is nearly 0.95. Because the dependent and independent variable are both log transformed, this coefficient indicates that a 1% increase in the prior year's rent would cause a 0.95% increase in the current period rent. When median income and the dummy variables are removed in equations (3-5), (3-6), and (3-7), the coefficient increases to 0.99, indicating that 1% increase in the prior year's rent would increase the current rent by

0.99%. It should be noted that in all six regressions where $RENT_{t-1}$ is included, it is significant at 99% confidence.

5.3 One-Year Lagged Trips

$TRIPS_{t-1}$also has a positive correlation as expected. This indicates that that prior year taxi activity can be used to predict current year rents. Interestingly, the number of pick-ups, which is used in equation (3-2) has less of an effect on RENT than either the drop-offs or the total of pick-ups and drop-offs, used in equations (3-3) and (3-4) respectively, while yielding a better model fit. This may suggest that pick-ups are not as influential as drop-offs are to rent changes. The coefficients obtained range from 0.003 to 0.010, with the combination of pick-ups and drop-offs having the highest coefficient and pick-ups having the lowest coefficient. As all values are log transformed, the coefficients can be interpreted to mean that a 1% change in pick-ups predicts a 0.003% change in RENT, a 1% change in drop-offs predicts a 0.005% change in RENT, and a 1% change in pick-ups and drop-offs predicts a 0.01% change in RENT.

In equations (3-5), (3-6), and (3-7), the coefficients decrease slightly, ranging from 0.002 to 0.009. In these equations, a 1% increase in trips corresponds to an increase between 0.002% and 0.009% in RENT, depending on the definition of trips.

In equations (3-8), (3-9), and (3-10), the coefficients increase dramatically, ranging from 0.039 to 0.144 with pick-ups having the lowest coefficient and pick-ups and drop-offs having the highest coefficient. In these equations, a 1% increase in trips corresponds to an increase between 0.039% and 0.144% in RENT, depending on the definition of trips.

In all cases, the $TRIPS_{t-1}$is found to be significant at 99% confidence.

5.4 One-Year Lagged Income

INC_{t-1}has a positive correlation as expected, however it's magnitude is lower than expected. The positive correlation indicates that prior year median income can be used to predict current year rent. In all three regressions utilizing INC_{t-1}, the same coefficient of 0.057 is obtained. The consistency of this result indicates that a 1% change in median income will correspond to a 0.057% increase in the next year's rent. As is true with RENT and TRIPS, the coefficient is significant at 99% confidence in all regressions in which it is present.

5.5 Dummy Year and Borough Variables

In equations (3-2), (3-3), and (3-4), the Year Dummy Variables have negligible coefficients and are not statistically significant. This indicates that year fixed effects are not useful in determining rent changes when utilizing rent, trips, and income as the other predictor variables. Surprisingly, the same can be said of the borough dummy variables.

5.6 Model Fit

The adjusted R^2 value of the regressions on equations (3-2), (3-3), and (3-4) are nearly 0.99, an unprecedented result. This means that the models can explain 99% of log rent values using log transformed time lagged rent, income and trips values. However, because of the log transformations, this is not surprising. Because rent is generally a change from the prior year's rent, the prior year's rent is very close to the next years, making the log transformed results not particularly useful.

To test this result further, the regression is performed again on equations (3-5), (3-6), and (3-7), which exclude income, due to possible autocorrelation with rent, and exclude the dummy variables, due to lack of statistical significance. The regression outputs are shown in Table 5-2. The model fits are nearly identical and adds confidence to the initial result. The coefficients on time lagged rent have increased, suggesting again that RENT is highly positively correlated to past rent.

Last, all variables other than trips are excluded from the regression according to equations (3-8), (3-9), and (3-10), because of the highly positive correlation between prior rent and current rent. The regression outputs are shown in Table 5-3. The adjusted R^2 values in this regression are much more reasonable and vary between 0.0097 an 0.147 with pick-ups as taxi activity providing the worst fit and drop-offs as taxi activity providing the best fit. This suggests that pick-ups are less predictive of rent, explaining its low coefficient value in equation (3-2). In all three models, the taxi activity is found significant at 99% confidence. This model indicates that a 1% increase in drop-offs above the borough and year mean, corresponds with a 0.07% increase in median rent.

5.7 Discussion

It was initially hypothesized that taxi activity could be used as a proxy for exogenous changes to neighborhoods, such as rising income and neighborhood quality. The regression results are highly significant and produced a well fit model. Thus, taxi trips can be thought of as a reasonable proxy for neighborhood changes that affect rents.

When $RENT_{t-1}$ is removed from the regression, the coefficients on $TRIPS_{t-1}$increase substantially, and the definition of trips as the number of drop-offs provides the best model fit. The increase in coefficients indicates that some of the explanatory power of time lagged trips are absorbed by time lagged rent. The best model fit being with drop-offs as taxi activity indicates that drop-offs might be the most accurate predictor of rent changes.

The effect of a 1 standard deviation positive shock in taxi drop-offs will result in a 0.009% - 0.154% higher rent the next year on average. The precise effect was not determined in this study, but further research should be conducted to arrive at a narrower estimation of the effect. However, this research has determined that there is a positive and significant correlation between taxi activity one year prior and the current median rent.

The median rent data in this study underestimates market rent because of the survey methods used in the American Community Survey. The rent information includes housing that does not respond to market forces, but rather legislative ones. It is likely that the effect on pure market rate residential rents will be greater because they move with the market demand, while rents subject to forms of rent control do not. Further study should be conducted using pure market rents to determine the predictive power of taxi activity on unregulated rents.

This study utilizes taxi data from the period when application based ride-hailing services were beginning to enter the market. The effect of exclusion of this data from the models may also cause the effects of taxi activity to be underestimated because the full amount of taxi activity is itself underestimated. Ride hailing data is treated as proprietary by its service providers and is typically not available for analysis. The demographics of ride hailing application users is not publicly available, but is known by the service providers. Information about the trips and users of ride hailing services would be necessary to opine on the possible effects of including the data.

All of the data analyzed in this study was collected during a period general rent increases in New York City. It is unknown if taxi activity would be a statistically significant predictor for rent changes in a downward period of the real estate cycle.

The time of day and time of year of taxi activity are ignored in this study. The possibility exists that trips taken during a certain time of day or during a certain time year act as a better predictor for rent than yearly totals of activity.

Despite the limitations on the data used in this study, there is strong evidence to support that taxi activity can predict the direction in which rents will move in a given neighborhood. Further research is necessary to fully understand the predictive power of taxi activity on residential rents.

	Dependent variable:		
		RENT	
	(3-2)	(3-3)	(3-4)
tilde{RENTt-1}	0.957***	0.953***	0.953***
	(0.007)	(0.007)	(0.007)
tilde{TRIPSt-1}	0.003***	0.005***	0.010***
	(0.0005)	(0.001)	(0.001)
tilde{INCt-1}	0.057***	0.057***	0.057***
	(0.009)	(0.009)	(0.010)
as.factor(Year)2012	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
as.factor(Year)2013	0.00001	0.001	0.001
	(0.002)	(0.002)	(0.002)
as.factor(Year)2014	0.0001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
as.factor(Year)2015	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
as.factor(Borough)Brooklyn	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
as.factor(Borough)Manhattan	-0.000	-0.000	-0.000
	(0.003)	(0.003)	(0.003)
as.factor(Borough)Queens	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)
as.factor(Borough)Staten Island	0.0001	0.00005	0.0001
	(0.004)	(0.003)	(0.003)
Constant	-0.00001	-0.00000	-0.00001
	(0.002)	(0.002)	(0.002)
Observations	905	935	935
R^2	0.989	0.988	0.988
Adjusted R^2	0.989	0.988	0.988
Residual Std. Error	0.023 (df = 893)	0.024 (df = 923)	0.024 (df = 923)
F Statistic	7,293.735*** (df = 11; 893)	7,008.357*** (df = 11; 923)	6,991.408*** (df = 11; 923)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5-1

	Dependent variable:		
	RENT		
	(3-5)	(3-6)	(3-7)
tilde{RENTt-1}	0.991***	0.988***	0.988***
	(0.004)	(0.004)	(0.004)
tilde{TRIPSt-1}	0.002***	0.004***	0.009***
	(0.0005)	(0.001)	(0.001)
Constant	0.00001	-0.00001	-0.00001
	(0.001)	(0.001)	(0.001)
Observations	905	935	935
R^2	0.989	0.988	0.988
Adjusted R^2	0.989	0.988	0.988
Residual Std. Error	0.024 (df = 902)	0.024 (df = 932)	0.024 (df = 932)
F Statistic	38,932.640*** (df = 2; 902)	37,428.990*** (df = 2; 932)	37,330.150*** (df = 2; 932)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5-2

	Dependent variable:		
	RENT		
	(3-2)	(3-3)	(3-4)
tilde{TRIPSt-1}	0.039***	0.070***	0.144***
	(0.004)	(0.005)	(0.012)
Constant	0.0002	-0.0004	-0.0004
	(0.007)	(0.007)	(0.007)
Observations	905	935	935
R^2	0.096	0.147	0.141
Adjusted R^2	0.095	0.147	0.140
Residual Std. Error	0.210 (df = 903)	0.201 (df = 933)	0.202 (df = 933)
F Statistic	95.378*** (df = 1; 903)	161.373*** (df = 1; 933)	152.995*** (df = 1; 933)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5-3

6. Conclusion

This study examined whether prior year taxi activity, along with prior year median rent and prior year median income, can be used to predict the median rent in a neighborhood. 188 neighborhoods in New York City were examined over the period of 2010-2015. Taxi activity was defined three ways: (i) the annual number or taxi pick-ups, (ii) the annual number of taxi drop-offs, and (iii) the annual number of pick-ups and drop-offs. A total of 1,466,234,991 pick-ups and drop-offs were located inside of the studied neighborhoods. Multivariate linear equations, in which all independent and dependent variables are log transformed and demeaned, were estimated for each definition of taxi activity.

The multivariate linear regression models find that taxi activity is a statistically significant determinant of residential rents at 99% confidence, in all definitions of taxi activity. The models indicate that the fixed effects of the borough in which a neighborhood is located and the year in which taxi activity occurs are not statistically significant to the result. When time lagged rent is included in the model, the adjusted R^2 value approaches 0.99. When rent is modeled solely as a function of taxi activity, the best fit model is the one which uses drop-offs as the indicator of taxi activity. This model has an adjusted R^2 value 0.147 and indicates that that each percent increase in drop-offs above the borough and year mean, corresponds with a 0.07% increase in median rent. A one standard deviation positive shock in the number of taxi drop-offs in a neighborhood corresponds with a median rent increase between 0.009% and 0.154% in the next year.

Past research on residential rents finds that median income and neighborhood quality are significant and positively correlated determinants of residential rent. When residential rents increase due to an influx of wealthier residents, it commonly referred to as gentrification. Gentrification occurs at the neighborhood level in stages. Early stage gentrification can be predicted by the over representation of certain business that employ younger workers. Later, consumption amenities arrive in these neighborhoods which may trigger the arrival of wealthier residents. Taxis users in New York City have been found to be younger and wealthier than the general population. This study speculates that taxi activity is a proxy for neighborhood quality and median income changes and can indicate the occurrence of gentrification and the increase in residential rents because of the demographic most likely to be taxi users. The regression outcomes support this speculation.

The research performed in this study is a first step towards understanding the predictive power of taxi activity on residential rents. Further research that arrives at a more precise coefficient for taxi activity, includes data from ride hailing services, uses pure market rents as the dependent variable, and considers the time of day or time of year of taxi activity is warranted.

7. References

Behrens, K., Boualam, B., Martin, J., & Mayneris, F. (2018). *Gentrification and pioneer businesses* (No. 13296). Retrieved from C.E.P.R. Discussion Papers website: https://ideas.repec.org/p/cpr/ceprdp/13296.html

Clay, P. L. (1979). *Neighborhood renewal: Middle-class resettlement and incumbent upgrading in American neighborhoods*. Lexington Books.

Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*, *22*(3), 433–452. https://doi.org/10.1016/0166-0462(92)90038-3

Gilderbloom, J. I., & Appelbaum, R. P. (1987). Toward a Sociology of Rent: Are Rental Housing Markets Competitive? *Social Problems*, *34*(3), 261–276. https://doi.org/10.2307/800766

Glass, R. (1964). *London; aspects of change* (University College, London. Centre for Urban Studies. Report ; no. 3). London: MacGibbon & Kee.

Guerrieri, V., Hartley, D., & Hurst, E. (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics*, *100*, 45–60. https://doi.org/10.1016/j.jpubeco.2013.02.001

Marks, D. (1984). The Effect of Rent Control on the Price of Rental Housing: An Hedonic Approach. *Land Economics*, *60*(1), 81–94. https://doi.org/10.2307/3146095

New York City Department of City Planning. (n.d.). Neighborhood Tabulation Areas (Formerly "Neighborhood Projection Areas"). Retrieved June 26, 2019, from https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-nynta.page

New York City Department of City Planning. (n.d.). NYC Population - American Community Survey. Retrieved June 26, 2019, from https://www1.nyc.gov/site/planning/data-maps/nyc-population/american-community-survey.page

New York City Taxi and Limousine Commission. (2014). *2014 Taxicab Fact Book*. Retrieved from https://www1.nyc.gov/assets/tlc/downloads/pdf/2014_tlc_factbook.pdf

New York City Taxi and Limousine Commission. (n.d.). TLC Trip Record Data. Retrieved June 26, 2019, from https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

NYU Furman Center. (2013). *State of New York City's Housing and Neighborhoods in 2013*. Retrieved from New York University website: http://furmancenter.org/files/sotc/SOC2013_HighRes.pdf

Renwick, T. (2011). *Geographic Adjustments of Supplemental Poverty Measure*

Thresholds: Using the American Community Survey Five-Year Data on Housing Costs. 32.

Rosen, S. Z. (1974). *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition*. https://doi.org/10.1086/260169

Sirmans, G., & D. Benjamin, J. (1991). Determinants of Market Rent. *Journal of Real Estate Research*, *6*, 357–380.