Pace University
DigitalCommons@Pace

Honors College Theses | Pforzheimer Honors College

5-2022

# Infrastructure Development and Gentrification: A Case Study of the 2017 Q Line Extension in New York City

Harrison Shoaf

Follow this and additional works at: https://digitalcommons.pace.edu/honorscollege_theses

Part of the Economics Commons

# INFRASTRUCTURE DEVELOPMENT AND GENTRIFICATION: A CASE STUDY OF THE 2017 Q LINE EXTENSION IN NEW YORK CITY

Harrison Shoaf

Kier Hanratty

Economics

May 3rd, 2022

May 16th, 2022

## Abstract

An examination of the 2017 Q Line subway extension in New York City and the potential causal relationship between its implementation and rental rates and gentrification in the surrounding area. Analysis of data covering the timeframe from 2007 to 2019 allows for utilization of OLS regression to determine if the area subject to the implementation experienced a change in rental rates and instigation of gentrification afterward compared to areas that were not subject to the implementation. Results indicate a decrease in rental rates (and by extent, no instigation of gentrification) in the area subject to the extension after it was implemented. The current recommendation is to continue developing infrastructure in pursuit of maximum economic growth/efficiency. Potential future research lies in the analysis of other infrastructure projects in other geographies utilizing similar methods as done here.

# Table of Contents

# List of Figures and Tables

**Section 1. Introduction**

Private housing has seen an exceptional rise in pricing in recent history, with some paralleling the rise to the financial crisis in 2008. As housing data for the first quarter of 2021 has become available, national media is noting the impact of this price level increase, specifically that it has been higher than at any point in the last 30 years (Badkar, 2021). Additionally, gentrification, a process of neighborhood change that entails economic change in a historically disinvested neighborhood by means of real estate investment and commercial development, can further exacerbate inequality in terms of both income levels and demographics. If the overarching goal is to stabilize/reduce rates in the housing market and/or reduce the impacts of gentrification, the question becomes one of what can be actively done to aid these two pursuits rather than doing nothing and letting the housing market/other external factors affect affairs as they will.

The difference between gentrification and simple economic growth is one of activity; if the quality of an area grows over time, assuming that growth is uniform, the population thereof should follow suit in terms of income level. However, in the case of gentrification, the change is much more sudden and active. Additionally, the population is typically “replaced” rather than “improved” in terms of overall welfare. Therefore, it is reasonable to ponder that increasing accessibility between affluent and poorer neighborhoods may lend itself to gentrification if the scale is not balanced, so to speak. Simply put, offering more affluent communities access to the amenities of poorer communities will allow them to take advantage of those amenities. However, the relationship isn’t exactly symmetric as those of a poorer upbringing may have a harder time utilizing the amenities available to affluent communities.

This distinction is being drawn because of a curious case of infrastructure development in New York City, specifically the 2017 Q Line Extension. Perhaps this extension contributed to an increase in overall rental rates due to the newly available proximity of various stations and the accessibility provided by them; in doing so, it may have contributed to gentrification. If poorer populations now have even less of an ability to afford housing, then more of it will be left to the more affluent communities. Once these affluent people begin living in this housing, they will, over time, improve it and the general community as far as their standards and financial means will allow them to. Of course, the financial means of the affluent are greater than that of the poor, and so, theoretically, the result would be gentrification and lesser availability of housing for poorer populations given that what once belonged to them is now too expensive for them to remain in.

**Section 2. Literature Review**

While infrastructure development has a relatively firm position in economic literature, the position of gentrification is more subjective. For the most part, existing literature on gentrification is incredibly balkanized and its exact definition or origin is not clearly defined (Slater 2003). Although some may think of it as such, gentrification is not a phenomenon that can easily be attributed to a single causal factor; rather, economic and cultural factors produce it in tandem (Ley 2003), and any arguments surrounding gentrification are contextual in both space and time (Slater 2003). One argument on gentrification concerns its relationship with overall neighborhood change. Initially, that change was viewed as a one-way process where the wealthy seldom revert their course and move backwards into the obsolete housing they are giving up. Gentrification undermines the dominant assumption that filtering of populations is a uni-directional downwards process where lower income groups move into progressively deteriorated buildings, and also challenges the assumption that the preference for space and low densities are far more important than accessibility to the central city.

It is important to note the differing discourses surrounding gentrification, its composition, and its potential causes (Hamnett 1984). Two potential factors in gentrification and, by extent, rental rates, are public policy and a structural shift of housing markets. Public policy has shifted from a welfare state based on direct public intervention into a reliance on free-market solutions, promoting partnership in areas that had previously relied solely on public funding, such as housing for low-income households (Hackworth 2007). A structural shift in the housing market may also facilitate gentrification as a result of increased availability of capital and credit (Hyra 2012).

Two notable discourses on gentrification find themselves divided by a physical border between the United States and Canada (Slater 2003). In the United States, a revanchist city discourse exists, one that argues that gentrification is the spatial manifestation of a race, class, and gender terror felt by middle-class white people who feel that their city has been stolen from them; in this framework, gentrification is an attempt to "retake" the city from the working class (Smith 1996). A theory on gentrification more commonly accepted in Canada is one of emancipatory discourse (Caulfield 1988), where gentrification is not pitting middle class against working class but against the repressive institutions of suburban life; it's much more about breaking free from the routine of placeless space and monofunctional instrumentality than about "taking over" the city (Caulfield 1989).

Considering this initial knowledge on gentrification, Slater decided to do a study on the gentrification of two different neighborhoods, South Parkdale (in the case of Toronto, Canada) and Lower Park Slope (in the case of New York City, USA). A notable conclusion of the analysis of gentrification in these areas is that gentrification would not be happening with tighter rent regulations enforced by higher levels of the government; without this practice, there is the displacing Tenant Protection Act in Canada and the "rent war" between the Federal Additional Compensation program and money-hungry landlords in the case of the US, both of which further exacerbate the gentrification problem. (Slater 2003).

Infrastructure development is an important force in rapid, sustained economic growth, with a positive correlation between economic development and quality of housing or access to amenities (Rao, Srinivasu 2013). The link between infrastructure development and economic growth is complicated; not only does it affect production and consumption directly, but it also creates many direct and indirect externalities (Rao, Srinivasu 2013). Additionally, infrastructure development affects output in two ways; one being a direct channel where infrastructure increases output by reducing the cost of intermediate goods and the other being an externality effect where infrastructure development results in higher human capital returns due to education, good quality health, and higher efficiency of human capital due to lower marginal depreciation of capital (Rao, Srinivasu 2013).

Countries beginning at a lower standard of living receive especially exponential growth and reductions in poverty rates due to new infrastructure development, especially when compared to more economically established countries (de la Fuente, 2004). Additionally, there's little indication that infrastructure development accounts for a significant portion of the negative impact on poverty rates and overall growth in any country (de la Fuente, 2004). It is important, however, to distinguish between regular economic growth/poverty reductions and gentrification as a phenomenon. Recall that gentrification entails economic change in a historically disinvested neighborhood by means of real estate investment and new, higher-income residents moving in. Gentrification takes a more active approach to growth and development, where higher-income residents begin to control the areas that previously were relegated to less affluent populations at a relatively rapid pace when compared to simple economic growth/development.

Although there is plenty of literature on infrastructure development and gentrification as separate phenomena, there seems to be little interconnecting the two. Recall that gentrification often has to do with rapid commercial and residential development in tandem. Thus, if it could be found that infrastructure development results in an increase in rental rates, then it can be asserted that actively developing massive infrastructure projects has a hand in gentrification. Perhaps the greater development and quality of amenities resulting from infrastructure development (Rao, Srinivasu 2013) may result in overall greater property values in the neighborhood in question or the city as a whole. These two factors would, in theory, raise property values and rents to such an extent that the historically less affluent populations that have lived in these places before may not have the means to live there any longer. Over time, this may gentrify the neighborhood commercially and residentially according to the revanchist theory surrounding gentrification (Smith 1996). The upper middle class may begin “taking over” the cities/neighborhoods that were once primarily composed of those of the working class or other less privileged demographics, such as racial minorities.

Even if these raising property values can result in involuntary evictions, there is still something to say about voluntary evacuations; people who sense or see gentrification occurring in their neighborhood may simply cut their losses and move to a more affordable area. Regardless of how voluntary this evacuation is, it still lends to gentrification, assuming that a more affluent, “middle-class” population takes control of the space that was once occupied by less fortunate groups. The way various socioeconomic classes move throughout various neighborhoods and areas is not quite symmetric. Of course, it is much easier for an affluent demographic to take advantage of the amenities and such offered by poorer communities/areas.

However, even from a strictly financial perspective, it is difficult for a less privileged demographic to utilize what is available to a more affluent one.

## METHODOLOGY

### Section 3. Pre-Estimation Discussion

As described previously, to determine a causal relationship between the implementation of the Q Line extension and gentrification, the model must consider several potential indicators of gentrification. While rapid changes in rental rates over time can be an indicator, it is not absolute in and of itself, given that gentrification contains both class and demographic components. As such, the model will include variables to both account for the components of gentrification and to control between the areas of interest for comparability.

The dependent variable in the model will be median rental rates over the timeframe of interest; 2007 to 2020. With rental rates being one of the more prevalent indicators of gentrification, including it as the dependent will allow for analysis of a causal relationship between the Q Line extension and gentrification, given that the model will also include other indicators of gentrification as independents. When considering these other factors, a significant change in rental rates may indicate gentrification rather than simple economic growth or changes/normal trends/fluctuations in the housing market. Data on median rental rates is stratified by neighborhood and retrieved from StreetEasy, a real estate search engine that helps potential buyers and renters with their search for housing and also provides a wealth of housing related data.

To account for the demographic component of gentrification, two variables will be separated by race, those being population count and graduation rates. Between the two of these, the reasoning behind the inclusion of population as a variable is more obvious; if there is a mass exodus or influx of a given demographic over a relatively short period of time, it is likely that a specific event or development is causing that change in the area's demographics. More specifically, if there is a significant increase in rental rates, this may be accompanied by an exodus of demographics that are disproportionately affected by poverty. Inversely, if a given neighborhood or area finds itself subject to an influx of such a population, rental rates may decrease over time so that housing can still be rented out and there is not a standstill in the market. Data on population count by race is collected from the United States Census Bureau, specifically the American Community Survey, table DP05.

The model also includes racial graduation rates to account for another side of gentrification's demographic component; propensity to succeed. To elaborate, a disproportionate change in the graduation rates of one race when compared to another may indicate that the race with the lower rate is disadvantaged in some way that prevents them from succeeding in school compared to students of other demographics. This can stem from a lack of comprehensive academic enrichment, resulting in only providing success opportunities to more privileged demographics, but it can also come from changes at home. Perhaps less privileged populations are less able to house themselves in a stable manner, leading to frequent moving that disrupts the development of students within these populations. Another possibility is that the problem is not the stability of housing, but rather medical issues afflicting students' family members, which may occur more often with less privileged populations with less access to quality healthcare. If the student needs to have a hand in caring for these family members, that may take away from their time to study and learn, and therefore their ability to graduate on time. Given that a number of these factors and situations can stem from and/or be exacerbated by gentrification, the inclusion of racial graduation rates is justified. Data on racial graduation rates is stratified by school district and collected from the New York City Department of Education. The data is most readily available on NYC Open Data, a website that collects numerous datasets specifically pertaining to New York City. Data pertaining to several other variables is retrieved from NYC Open Data as well; these are noted as necessary.

An additional variable that can account for the class component of gentrification is crime incident count. To clarify, in order for the development caused by gentrification to maintain its hold, crime must drop and/or stabilize, as an increase in crime levels would give both potential residents and those engaging in commercial development less of an incentive to establish themselves in the area that is being gentrified. Pertaining to the model and discussion thereof, a significant decrease in crime post-implementation would provide additional support that the implementation did instigate or exacerbate gentrification. Data on crime incident count is stratified by police precinct and pertains to the incident count of the seven major felonies as defined by the New York Police Department, those being murder/non-legal manslaughter, rape, robbery, felony assault, burglary, grand larceny, and grand larceny of a motor vehicle. The total count of these incidents is the value recorded for each year/precinct. Data on these incident counts is collected from the New York Police Department.

Controls utilized in the model include median household income, median age, and mean family size, with all of these stratified by congressional district. The model utilizes these controls to further ensure comparability between the areas of interest. Each of these three factors can affect rental rates in a given area.

The median household income of a given area plays a role in where rental rates begin as well as the range that they tend to sit in. As is true with any market, housing can only be sold, or in this case, rented, to a reasonable extent if its typical pricing coincides with what most consumers are willing and able to pay; raising rates well above most consumers' means will result in a surplus that cannot be depleted until rates come back down, while tanking rates will eventually cause a severe shortage once the supply of housing has been rented out. Including median household income allows the model to control for comparability between each of the areas of interest.

Including median age allows the model to control for the effects that the age of a population may have on rental rates. Those of a younger demographic may be less likely to selectively bargain or search for the absolute most ideal housing in terms of both quality and price; they may be more inclined to take what they can get even if the pricing of what they get isn't exactly "fair" for what they are receiving. Conversely, while older populations may have more money to spend, moving into a new home is often a matter of preference rather than necessity for people of these ages. Thus, they are less inclined to go through with purchasing or renting homes with exorbitant costs than comparatively younger demographics. Including age as a control ensures that the effect of the implementation found by the model is truly due to the implementation and not the effect that age may have on rental rates in a given area.

Finally, there is mean family size. The most obvious way in which this can affect rental rates is that larger families need housing with more space and/or rooms to accommodate everyone who is living with them. Thus, if most rentals in a given area are done with relatively large families, that may increase the overall rate, and so inclusion of family size as a control for rental rates allows for further corroboration that the change in rental rates is indeed a result of the implementation rather than changes in the typical family size in the area.

Data on all three of these control variables is, again, stratified by congressional district and retrieved from the United States Census Bureau, specifically the American Community Survey, tables DP05 (in the case of median age), S1101 (in the case of mean family size), and S1903 (in the case of median household income).

As far as shortcomings when it comes to data collection and compilation go, there are two that come to mind: lack of perfectly identical stratification available across all variables and missing values in some years for some variables, most notably at either the very beginning or very end of the timeframe of interest. For example, in the case of variables collected from the Census Bureau, the shortcoming lies in the fact that the American Community Survey, the most comprehensive source available for the purposes of this study, only goes back to the year 2010. In the case of the unequal stratification problem, a step taken to rectify it was the utilization of several geography maps of each stratification[1] to ensure geographic proximity/equality between stratifications to the furthest extent possible. Additionally, to corroborate these equivalencies, a data sheet of crosswalks[2] was utilized to ensure technical equality in addition to geographic equality provided by the geographic maps.

Collective use of all these stratifications allowed for specification of several different geographic areas to compare when it came time to run the regressions necessary to ascertain a causal effect on rental rates and gentrification from the Q line extension. A caveat of the study was that the data could not include a great number of areas due to the threat of great heteroskedasticity, which only got worse as the data began to include more and more areas. However, the data still includes a number of them, defined as follows:

Area 1: School District 2, Upper East Side (79$^{th}$ - 96$^{th}$) 12th Congressional District, 19th Precinct

Area 2: School District 25, Flushing (North), 6th Congressional District, 109th Precinct

---

[1] Stratification maps were provided by GovTrack (Congressional District), NYPD (Police Precinct), and the NYC Department of City Planning (Neighborhood and School District)

[2] Varun Adibhatla, NYC-LocalGeo-Crosswalk (2017)

Area 3: School District 13, Brooklyn Heights, 7th Congressional District, 84th Precinct

Area 4: School District 10, Riverdale, 16th Congressional District, 50th Precinct

Area 5: School District 31, Silver Lake, 11th Congressional District, 120th Precinct

Area 6: School District 3, Upper West Side ($79^{th}$ - $96^{th}$), 10th Congressional District, $20^{th}$ Precinct

Of these areas, Area 1 pertains to the area subject to the Q Line Implementation

A previous iteration of this study contained only areas 1-5, which resulted in no heteroskedasticity whatsoever. Even with just the inclusion of this $6^{th}$ area, heteroskedasticity appears (see Appendix, Figure 3). Since the dependent variable as well as many others are already logged and the other variables in the model cannot be redefined in any meaningful way, the most that can be done to alleviate this problem is application of robust standard errors. These robust standard errors are applied for all regressions after the first. Inclusion of additional areas in the data would have only further exacerbated this problem, and so that is why, unfortunately, inclusion of a massive number of areas is not a reasonable option.

The following figure is a summary statistics table concerning all variables included in the regressions ran. Most variables that appear in the actual dataset constructed from the data gathered are used in the regressions, but some, such as year or area, are for organizational purposes only and thus are not used:

**Figure 1: Summary Statistics Table for Variables Used in Regression**

Descriptive Summary of Statistics

| | lnrent | lnasian | lnblack | lnhisp | lnwhite | agrad | bgrad | hgrad | wgrad | lncrime | lnincome | age | fam | treat | post |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mean | 7.786191 | 11.45609 | 11.21454 | 11.93916 | 12.78537 | 82.08889 | 52.50694 | 50.32361 | 77.80278 | 7.187174 | 11.06065 | 36.94167 | 2.613667 | .1666667 | .2857143 |
| sd | .273558 | .7728975 | .9094774 | .4747091 | .4186038 | 8.665937 | 8.381188 | 8.333142 | 9.424077 | .3437622 | .331511 | 3.088869 | .3549408 | .3749163 | .4544672 |
| p25 | 7.56294 | 11.371 | 10.43408 | 11.53288 | 12.70695 | 77 | 46.35 | 43.45 | 71.15 | 6.881923 | 10.92018 | 34.65 | 2.325 | 0 | 0 |
| p75 | 8.034105 | 11.8419 | 12.33792 | 12.21802 | 13.14194 | 87.75 | 58.75 | 56.75 | 85.15 | 7.553285 | 11.30541 | 39.75 | 2.8 | 0 | 1 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 64 | 60 | 60 | 60 | 60 | 72 | 72 | 72 | 72 | 84 | 60 | 60 | 60 | 84 | 84 |
| *N* | 84 | | | | | | | | | | | | | | |

Data Retrieved from StreetEasy, US Census Bureau, New York Police Department, NYC Department of Education, NYC Department of Finance

To reiterate, the hypothesis of this case study is that the Q Line extension resulted in an increase in rental rates (and therefore gentrification) due to the greater accessibility that the additional stations bring, thereby increasing property values in the process of the implementation of the line.

**Section 4. Estimation**

The baseline model utilized to determine the potential causal effect of the Q Line's implementation on rental rates and gentrification is expressed in the following equation:

$$\begin{aligned} lnrent_{c,t} = {} & \beta_0 + \beta_1 lnasian_{c,t} + \beta_2 lnblack_{c,t} + \beta_3 lnhisp_{c,t} + \beta_4 lnwhite_{c,t} + \beta_5 Agrad_{c,t} \\ & + \beta_6 Bgrad_{c,t} + \beta_7 Hgrad_{c,t} + \beta_8 Wgrad_{c,t} + \beta_9 lncrime_{c,t} \\ & + \beta_{10} lnincome_{c,t} + \beta_{11} Famsize_{c,t} + \beta_{12} Age_{c,t} + \beta_{13} Treat_{c,t} + \beta_{14} Post_{c,t} \\ & + \beta_{15} TreatXPost_{c,t} + e_{c,t} \end{aligned}$$

Where:

- lnrent is the natural log of median rental rates
- lnasian, lnblack, lnhisp, and lnwhite are the natural log of the counts of the Asian, Black, Hispanic, and White populations, respectively
- Agrad, Bgrad, Hgrad, and Wgrad are the graduation rates for Asian, Black, Hispanic, and White students respectively
- lncrime is the natural log of the incident count of the seven major felonies as defined by NYPD
- lnincome is the natural log of median household income
- Famsize is mean family size
- Age is median age

- Treat is a binary variable denoting whether or not an area is subject to the treatment (the implementation of the extension)
- Post is a binary variable denoting whether the time is pre or post treatment
- TreatXPost is an interaction between the previous two variables that indicates any change in rental rates post treatment in the area that was subject to the treatment

An additional model with additional terms is expressed in the following equation:

$$\begin{aligned}lnrent_{c,t} = \ &\beta_0 + \beta_1 lnasian_{c,t} + \beta_2 lnblack_{c,t} + \beta_3 lnhisp_{c,t} + \beta_4 lnwhite_{c,t} + \beta_5 Agrad_{c,t} \\ &+ \beta_6 Bgrad_{c,t} + \beta_7 Hgrad_{c,t} + \beta_8 Wgrad_{c,t} + \beta_9 lncrime_{c,t} \\ &+ \beta_{10} lnincome_{c,t} + \beta_{11} Famsize_{c,t} + \beta_{12} Age_{c,t} + \beta_{13} Treat_{c,t} + \beta_{14} Post_{c,t} \\ &+ \beta_{15} TreatXPost_{c,t} + \beta_{16} lnincomeXlnasian_{c,t} + \beta_{17} lnincomeXlnblack_{c,t} \\ &+ \beta_{18} lnincomeXlnhisp_{c,t} + \beta_{19} lnincomeXlnwhite_{c,t} + e_{c,t}\end{aligned}$$

Where lnincomeXlnasian, lnincomeXlnblack, lnincomeXlnhisp, and lnincomeXlnwhite are interactions between the natural log of income and the natural log of the counts of the Asian, Black, Hispanic, and White populations, respectively.

The most noticeable aspect of these models is the logging of numerous variables. This is because the data collected for many variables came in the form of a raw count rather than a percentage. While, in theory, an option is to keep all these variables linear and offer a technically correct interpretation of them within the context of the model, doing so makes little practical sense. Logging these terms provides the means to interpret much of the results in terms of percentages, which is much more practical when it comes to both illustrating the magnitudes of the effects each variable has on rental rates and for recommendations of policy/other action.

The final model includes interactions between income and racial population to account for potential racial discrepancies in income distribution. This may provide further explanation or control further for why rental rates are at a certain level before even considering the line's implementation; if a given area's population is predominantly of under-privileged demographics, that may have a significant effect on income levels in that area, and thus rental rates. Necessary diagnostics throughout the construction of the model and the running of regressions included a Breusch-Pagan/Cook-Weisberg test for heteroskedasticity on the baseline model, several F-Tests for joint insignificance for each set of the racially based variables at each iteration of the model, as well as the application of robust standard errors in both iterations of the model after the first.

## RESULTS

### Section 5. Post-Estimation Discussion

Running the regressions of the base model, the base model with robust standard errors applied, and the model with additional interaction terms applied gives the following results:

**Figure 2: Regression Results (Simplified; see Appendix, Figure 2.1 for full table)**

Base Model, Robust Standard Errors Added, Race Interactions Added

| | Base | RSE | Interactions |
|---|---|---|---|
| Asian Pop. | 0.20*** | 0.20*** | 0.06 |
| | (0.03) | (0.04) | (1.53) |
| Black Pop. | 0.07* | 0.07 | 3.00 |
| | (0.03) | (0.04) | (1.49) |
| Hisp. Pop. | 0.11* | 0.11** | 1.65 |
| | (0.04) | (0.04) | (1.69) |
| White Pop. | -0.07 | -0.07 | 5.65 |
| | (0.05) | (0.06) | (3.33) |
| Crime | -0.08 | -0.08 | -0.08 |
| | (0.09) | (0.09) | (0.09) |
| Asian Grad. | -0.00 | -0.00 | -0.01* |
| | (0.00) | (0.00) | (0.00) |

| | | | |
|---|---|---|---|
| Black Grad. | 0.01** | 0.01** | 0.01** |
| | (0.00) | (0.00) | (0.00) |
| Hisp. Grad | - 0.00 | -0.00 | -0.00 |
| | (0.00) | (0.00) | (0.00) |
| White Grad | 0.02*** | 0.02*** | 0.02*** |
| | (0.00) | (0.00) | (0.00) |
| Income | -0.15 | -0.15 | 11.07 |
| | (0.09) | (0.11) | (5.55) |
| Age | -0.02* | -0.02* | -0.01 |
| | (0.01) | (0.01) | (0.01) |
| Family Size | -0.53*** | -0.53*** | -0.50*** |
| | (0.09) | (0.11) | (0.12) |
| Treat | 0.14* | 0.14* | 0.12 |
| | (0.07) | (0.06) | (0.07) |
| Post | 0.02 | 0.02 | 0.02 |
| | (0.03) | (0.03) | (0.03) |
| TreatXPost | -0.09 | -0.09* | -0.10* |
| | (0.05) | (0.03) | (0.04) |
| IncomeXAsian | | | 0.01 |
| | | | (0.14) |
| IncomeXBlack | | | -0.27 |
| | | | (0.13) |
| IncomeXHisp | | | -0.14 |
| | | | (0.16) |
| IncomeXWhite | | | -0.52 |
| | | | (0.30) |

| | | | |
|---|---|---|---|
| _cons | 6.94*** | 6.94*** | -116.27 |
| | (1.41) | (1.37) | (61.59) |
| N | 52 | 52 | 52 |
| R-sq | 0.972 | 0.972 | 0.975 |
| adj. R-sq | 0.960 | 0.960 | 0.960 |

Note: Racial graduation rates, age, family size, and treat and post (the interaction as well), are NOT logged; all other variables are, including rental rates.

Standard errors in parentheses

* p<.05, ** p<.01, *** p<.001

**DISCUSSION**

The most notable coefficient to interpret from the regressions is that of the interaction between treat and post, quantifying the change in rental rates post treatment in the area the treatment was implemented. When it comes to the base model with robust standard errors applied, results indicate that, post implementation, there was a 9% decrease in the median rental rate in the area subject to the treatment. Although this effect becomes significant with the application of robust standard errors, we must still be cautious given that the model was subject to heteroskedasticity (see Appendix, Figure 3) in the first place, even though we have done all we can to rectify and/or weaken the impact of the problem.

Next, there is the question of the coefficients surrounding racial population demographics, namely Asian, Black, Hispanic, and White populations, with all coefficients except that of White population being significant at a reasonable level. Results from the base model with RSE indicate that a 1% increase in the Asian population results in a 0.20% increase in the median rental rate. Results also indicate that a 1% change in the White population results in an 0.07% decrease in the median rental rate. Regarding changes in the population of other races, for every 1% change in the Black population there is a 0.07% increase in the median rental rate, and every 1% increase in the Hispanic population results in a 0.11% increase. A diagnostic performed to further ensure the significance of racial population variables is an F-Test for joint significance. The tests did not find joint insignificance in the case of the variables for racial population in any iteration of the model except for when interaction terms between race and income were applied (see Appendix, Figure 4).

Concerning crime count, there is a 0.08% decrease in rental rates for every 1% increase in the number of incidents. With racial graduation rates, results indicate small changes in the median rental rate for every additional percentage point increase of the graduation rate of each race, those being a no change in the median rental rate in the case of Asian graduation rates, a 1% increase in the case of Black graduation rates, no change in the case of Hispanic graduation rates, and a 2% increase in the case of White graduation rates. Finally, there is income, where there is a 0.15% decrease in the median rental rate for every 1% increase in median household income.

As with racial population demographics, it was also necessary to conduct F-Tests on the graduation rate variables; the tests did not find joint insignificance for racial graduation rates in any iteration of the model (see Appendix, Figure 5).

The last two variables seen within the baseline model are age and family size, with both being significant at reasonable levels. In the case of age, results indicate that a 1 year increase in median age results in a 2% decrease in the median rental rate, while, according to the output of the regression, a 1 person increase in mean family size results in a 53% decrease in the median rental rate.

The third and final model detailed in the above regression table includes the entirety of the baseline model with the addition of several interaction terms between income and each of the races considered by the data; robust standard errors are also applied to this model. Firstly, with the interaction between treat and post in this third model, results indicate a 10% decrease in the median rental rate in the area where the treatment was applied post treatment.

Concerning the interaction terms between income and race themselves, results indicate that, holding income constant, a 1% increase in the Asian population results in a 0.01% increase in the median rental rate, while a 1% increase in the Black population results in a 0.27% decrease, a 1% increase in the Hispanic population results in a 0.14% decrease, and a 1% increase in the White population results in a 0.52% decrease. Even with these coefficients in mind, it is important to note that, when conducting a joint F-Test, the interaction terms themselves were found to be jointly insignificant, that is, we must approach the results on these terms with skepticism and cannot blindly accept perfect validity at face value (see Appendix, Figure 6).

Concerning racial graduation rates, in the model including the additional interaction terms, a 1 percentage point increase in the Asian graduation rate results in a 1% decrease in the median rental rate, a 1 percentage point increase in the Black graduation rate results in a 1% increase, and a 1 percentage point increase in the White graduation rate results in a 2% increase. As with the previous iteration of the model, changes in the Hispanic graduation rate do not result in any change in the median rental rate.

With crime count, the third model indicates that a 1% increase in the incident count results in a 0.08% decrease in the median rental rate. The result for income changed dramatically between the base model and the model with the additional interaction terms; results in the third model indicate that for every 1% increase in median household income, there is a 11.07% increase in the median rental rate. The effect of age on rental rates is about the same as in the base model, that is, a 1% decrease in the median rental rate for every additional year increase in median age. Finally, the magnitude of the effect that family size has on rental rates shrinks somewhat in the third model, with results indicating that for every 1 person increase in family size, the median rental rate decreases by 50%.

Some of the results found here are confusing, notably that of the interaction between the variables for treatment and post, which did indicate a change in rental rates as a result of the Q Line's implementation, although the change was a significant decrease in median rental rates. The confusion comes from the results contradicting the initial hypothesis, namely that rental rates would increase as a result of the greater accessibility (and thus property value) the additional stations would offer. Perhaps accessibility and/or interconnectedness is not as strong of a major factor in determining the actual value of property as initially thought, although that still seems odd given that people would have a higher demand for housing that makes them more interconnected with the world around them, so to speak.

The results surrounding changes in the racial demographics of the population are curious as well. With all racial minorities, in the base model there is an increase in rental rates whenever an increase in any of these populations occurs; a potential explanation could stem from attempts by those who deal in real estate to limit the influx of racial minorities into a given area by increasing rates as more come in, although this explanation would require further research in order to verify and test whether this effect occurs.

With crime, it is also curious that, according to any of the models, marginal increases in crime do not have a large impact on rental rates, although the direction of the effect is reasonable considering that areas with higher rates of crime tend to have less expensive housing. Perhaps the difference in crime needs to be much wider in order to see a noticeable change in rental rates. This mostly makes sense; while a small uptick in crime may be cause for concern, it won't necessarily result in a massive decrease in rental rates or property values, but a massive spike in crime may socially mark an area as unsafe, and so a majority of residents would flee until the only people left to buy housing there are those who are financially disadvantaged and thus don't have many other options for buying housing, eventually bringing rates down in the area with significant crime.

Although graduation rates were not expected to have huge, sweeping effects on rental rates, results overall indicate that marginal changes in these rates result in relatively minor changes in rental rates at best. Perhaps changes in graduation rates are not as strong of an indicator of the financial side of gentrification as initially thought, even if it may be an indicator of the demographic component.

Concerning income, the direction of the effect in the first two model causes some confusion; as consumers gain more and more income, they can afford more expensive housing, and thus if the general income of a population vastly increases, one would expect rental rates to follow suit. However, the magnitude and direction noted in the third iteration of the model (11.07% increase in median rate for every 1% increase in median household income), seems too large to be reasonable. The results on family size make some sense considering that higher income people tend to have fewer children, thus it would make sense that homes get more expensive as family size marginally decreases. The results on age make sense as well; many older residents are more adept with managing their money than those who are younger, and realtors are less able to get away with raising rates above what the property is actually worth, as older residents may know more about property values and/or be less desperate for housing compared to especially younger residents who may not be able to be choosy when it comes to their housing options due to lack of a significant amount of savings or a more urgent need to find housing even if its pricing is not ideal.

Finally, there are the interactions between income and race, which were found to be jointly insignificant. Generally, it is understood that if an interaction term turns out to be statistically insignificant, then it should not be included in the model. Because all interactions were found to be insignificant, it is clear that the superior model in terms of finding accurate, significant results is the baseline model with robust standard errors. However, even if the interaction terms had been jointly significant, making the model with interaction terms superior, the conclusions and policy recommendations drawn would hardly change in a qualitative sense, as the effect of the implementation (indicated by TreatXPost) does not change in direction and only somewhat changes in magnitude regardless of which iteration of the model is considered.

When it comes to the lack of statistical significance with many variables in the model, especially with interactions between race and income applied, this can stem from two factors, namely the sample size and the size of the standard errors. Concerning sample size; due to problems with heteroskedasticity it was difficult to incorporate a large number of areas without exacerbating that problem further; that combined with the relatively small timeframe (2007-2019) led to a smaller sample size than what may have been ideal. A larger sample size may have helped find results that are statistically significant more often than in the models seen here. With the standard errors, there is a similar story; some of these errors are of notable size given the coefficients they correspond to. This remains the case even with the application of robust standard errors and additional regressors.

Several caveats of the study have been described in previous sections. The first of these lies in the collection of the data in terms of both its stratification and its being comprehensive. By their nature, many of the variables utilized in this model were only available at the exact same stratifications at general levels, such as county or state. More specific stratifications were available for each variable, but not all of these were equal to each of the others in the model. Thus, use of geographic maps and crosswalks was necessary to equalize the data as much as possible, although the lack of absolute perfect equality may have contributed to the confusion found in the results of the regressions.

The second caveat lies in the number of areas as well as the appearance of heteroskedasticity. The ideal scenario would have been to include a large number of areas to further ensure the legitimacy of results. However, in pursuit of that goal, the heteroskedasticity problem only got worse. As detailed previously, including more and more areas led to a greater and greater propensity to heteroskedasticity; inclusion of even more areas would have made the problem even harder to solve than it was in this instance. If provided more time, the most direct solution to this problem would have been to innovatively redefine variables in the model and either transform or recollect data on those variables under those new definitions, although it is not clear how we could further go about redefining of variables, given that many are already logged. For the variables that are not logged, it makes no sense to do so to those variables, and, as with those that had to be logged, it is unclear how these linear variables can be redefined in a way that still allows for practical application of results.

A final caveat is the inability to include vacancy rates as a control variable in the model. Inclusion of vacancy rates as a control further and improve accuracy of the results, but unfortunately, data that was both comprehensive enough of the timeframe of interest and reported at a specific enough level to include in the model was not readily available despite efforts to search for it both in NYC Open Data and the Census Bureau; more general search efforts also did not locate the data necessary to include vacancy rates as a control.

**Section 6. Conclusion**

Based on the results seen here, there is no indication that the implementation of the Q Line instigated and/or exacerbated gentrification and/or increases in rental rates; if anything, results indicate a decrease in rental rates as a result of the implementation. With this in mind, the best recommendation for action drawn from these results is, if the goal is to reduce rental rates and/or gentrification, to not halt infrastructure development in the effort to do so. Because, according to these results, there is not a major increase in rental rates or gentrification as a result of the Q Line extension, it is reasonable to say that development and maintenance of infrastructure should continue to the extent necessary to promote maximum economic growth and efficiency; this can be done without great fear of an increase in rental rates or instigation of gentrification.

Potential avenues for future research mostly center around creating similar models pertaining to other instances of great infrastructure development. The lack of increase in rates seen in the case of the Q Line is curious; it would be necessary to verify that this lack of increase remains the case in most if not all other situations. More specifically, it would be best to verify if this lack of increase remains so regardless of the type of infrastructure project, the geographical location of the project, or other potential confounding factors. Ideally, the additional research would consider a development outside the realm of public transit, implemented in an area geographically separate from New York City. Seeing if this lack of increase is an isolated instance or not would determine if the recommendations for action given above would hold true universally. For the time being, however, there seems to be little reason to slow infrastructure development and maintenance in New York City if stabilizing rental rates and reducing the impacts of gentrification is a priority.

Increases in rental rates and gentrification are two phenomena that, if not managed, could make housing nearly unaffordable and/or drive already under-privileged demographics further into the ground in terms of their standard of living. The model constructed here aimed to determine if there is a causal relationship between infrastructure development and gentrification by analyzing the case of the 2017 Q Line Extension in New York City, hypothesizing that the improved accessibility offered by the station would significantly increase rental rates and thus instigate gentrification. Succinctly, the results of the economic models constructed to estimate this effect indicate either a decrease in rental rates in the surrounding area after the implementation of the line. Other variables within the model did see statistically significant effects on rental rates, but the implementation itself did not exacerbate either of these two phenomena.

**REFERENCES**


Badkar M, (2021, June 29th) "US home prices rise at fastest pace in more than 30 years", *Financial Times*, https://www.ft.com/content/88b4102f-a6d1-473a-8920-5e98b45661ae

Caulfield J, 1988, "Canadian Urban Reform and Local Conditions" *International Journal of Urban and Regional Research 12* 477 - 484.

Caulfield J, 1989, "Gentrification and Desire" *Canadian Review of Sociology and Anthropology 26* 617 – 632

De la Fuente, A. and A. Estache, (2004). "Infrastructure Productivity and Growth: A Quick Survey" Washington DC., WBIGF, mimeo, 2004.

GovTrack: Congressional Districts Map; https://www.govtrack.us/congress/members/map

Hackworth, Jason. 2007. The Neoliberal City: Governance, Ideology, and Development in American Urbanism. Ithaca, NY: Cornell University Press.

Hamnett, C. (1984) "Gentrification and residential location theory: a review and assessment", in Herbert, D. T. and Johnston, R. J. (eds) *Geography and the urban environment. Progress in Research and Applications, Vol. 6* (John Wiley, London)

Hyra, Derek S. 2012. "Conceptualizing the New Urban Renewal: Comparing the Past to the Present" *Urban Affairs Review 48*, Vol. 4, p. 498–527.

Ley D, 2003, "Artists, Aestheticisation and the Field of Gentrification" *Urban Studies 40* 2527 – 2544

New York Police Department (NYPD): Historical Crime Data by Precinct, Citywide Seven Major Felony Offenses, 2000-2020 https://www1.nyc.gov/site/nypd/stats/crime-statistics/historical.page

New York Police Department (NYPD): Precinct/Sector Map; retrieved from NYPD website; https://www1.nyc.gov/site/nypd/bureaus/patrol/find-your-precinct.page

NYC Department of City Planning (DCP): NYC Neighborhoods Map; https://www1.nyc.gov/site/planning/data-maps/city-neighborhoods.page#nycmap

NYC Department of City Planning (DCP): School Districts, Updated November 24th, 2021; retrieved from NYC Open Data

NYC Department of Education (DOE): 2017-2018 Graduation Outcomes District, retrieved from NYC Open Data; https://data.cityofnewyork.us/Education/2017-2018-Graduation-Outcomes-District/qvdv-dru4

Rao, P. Srinivasu, B, 2013, "Infrastructure Development and Economic Growth: Prospects and Perspective" *Journal of Business Management and Social Sciences Research 2* 81-91

Slater T, 2003, "North American Gentrification? Revanchist and Emancipatory Perspectives Explored" *Environment and Planning A 36*, 1191-1213

Smith N, 1996, "The New Urban Frontier: Gentrification and the Revanchist City" (Routledge, London)

Stoecker R, Vakil A, 2000, "States, Cultures and Community Organising: Two Tales of Two Neighborhoods" *Journal of Urban Affairs 22* 439 – 458

StreetEasy Data Dashboard, 2021: Median Asking Rent by Neighborhood; https://streeteasy.com/blog/data-dashboard/true?agg=Total&metric=Inventory&type=Sales&bedrooms=Any%20Bedrooms&property=Any%20Property%20Type&minDate=2010-01-01&maxDate=2022-01-01&area=Flatiron,Brooklyn%20Heights

United States Census Bureau, American Community Survey (Table DP050), ACS Demographic and Housing Estimates, 2010-2019;

https://data.census.gov/cedsci/table?q=kings%20county%20population&g=500XX00US3606,3607,3616,3611,3612&tid=ACSDP1Y2019.DP05&hidePreview=true&moe=false

United States Census Bureau, American Community Survey (Table S1101), Households and Families, 2010-2019: https://data.census.gov/cedsci/table?q=family%20size%20american%20community%20survey&g=500XX00US3606,3607,3616,3611,3612&tid=ACSST1Y2019.S1101&hidePreview=true

United States Census Bureau, American Community Survey (Table S1903), Median Income in the Past 12 Months, 2010-2019: https://data.census.gov/cedsci/table?g=500XX00US3606,3607,3616,3611,3612,3601&tid=ACSST1Y2019.S1903&hidePreview=true

Varun Adibhatla, NYC-LocalGeo-Crosswalk (2017), retrieved from GitHub; https://github.com/vr00n/NYC-LocalGeo-CrossWalk

**APPENDIX**

**Figure 2.1: Full Regression Table**

Base Model, Robust Standard Errors Added, Race Interactions Added

| | (1) | (2) | (3) |
|---|---|---|---|
| | est1 | est2 | est3 |
| lnasian | 0.20*** | 0.20*** | 0.06 |
| | (0.03) | (0.04) | (1.53) |
| lnblack | 0.07* | 0.07 | 3.00 |
| | (0.03) | (0.04) | (1.49) |
| lnhisp | 0.11* | 0.11** | 1.65 |
| | (0.04) | (0.04) | (1.69) |
| lnwhite | -0.07 | -0.07 | 5.65 |
| | (0.05) | (0.06) | (3.33) |
| lncrime | -0.08 | -0.08 | -0.08 |
| | (0.09) | (0.09) | (0.09) |
| agrad | -0.00 | -0.00 | -0.01* |
| | (0.00) | (0.00) | (0.00) |

| | | | |
|---|---|---|---|
| bgrad | 0.01** | 0.01** | 0.01** |
| | (0.00) | (0.00) | (0.00) |
| hgrad | -0.00 | -0.00 | -0.00 |
| | (0.00) | (0.00) | (0.00) |
| wgrad | 0.02*** | 0.02*** | 0.02*** |
| | (0.00) | (0.00) | (0.00) |
| lnincome | -0.15 | -0.15 | 11.07 |
| | (0.09) | (0.11) | (5.55) |
| age | -0.02* | -0.02* | -0.01 |
| | (0.01) | (0.01) | (0.01) |
| fam | -0.53*** | -0.53*** | -0.50*** |
| | (0.09) | (0.11) | (0.12) |
| 1.treat | 0.14* | 0.14* | 0.12 |
| | (0.07) | (0.06) | (0.07) |
| 1.post | 0.02 | 0.02 | 0.02 |
| | (0.03) | (0.03) | (0.03) |

| | | | |
|---|---|---|---|
| 1.treat#1.~t | -0.09 | -0.09* | -0.10* |
| | (0.05) | (0.03) | (0.04) |
| c.lnincome~n | | 0.01 | |
| | | (0.14) | |
| c.lnincome~k | | -0.27 | |
| | | (0.13) | |
| c.lnincome~p | | -0.14 | |
| | | (0.16) | |
| c.lnincome~e | | -0.52 | |
| | | (0.30) | |
| _cons | 6.94*** | 6.94*** | -116.27 |
| | (1.41) | (1.37) | (61.59) |
| N | 52 | 52 | 52 |
| R-sq | 0.972 | 0.972 | 0.975 |
| adj. R-sq | 0.960 | 0.960 | 0.960 |

Standard errors in parentheses

* $p<.05$, ** $p<.01$, *** $p<.001$

**Figure 3: Breusch-Pagan / Cook-Weisberg Test for Heteroskedasticity, Baseline Model**

Ho: Constant variance

Variables: fitted values of lnrent

chi2(1) = 2.97

Prob > chi2 = 0.0850

**Figure 4: F-Tests (Racial Population (Baseline Model, Robust SE Applied, Interaction Terms)**

Baseline Model:

( 1) lnasian = 0

( 2) lnblack = 0

( 3) lnhisp = 0

( 4) lnwhite = 0

F( 4, 36) = 24.96

Prob > F = 0.0000

Baseline Model with Robust SE Applied:

( 1) lnasian = 0

( 2) lnblack = 0

( 3) lnhisp = 0

( 4) lnwhite = 0

F( 4, 36) = 22.23

Prob > F = 0.0000

Interaction Terms Applied:

( 1) lnasian = 0

( 2) lnblack = 0

( 3) lnhisp = 0

( 4) lnwhite = 0

F( 4, 32) = 1.69

Prob > F = 0.1760

**Figure 5: F-Tests (Racial Graduation Rates; Baseline Model, Robust SE, Interaction Terms)**

Baseline Model:

( 1) agrad = 0

( 2) bgrad = 0

( 3) hgrad = 0

( 4) wgrad = 0

F( 4, 36) = 15.34

Prob > F = 0.0000

Robust SE Applied:

( 1) agrad = 0

( 2) bgrad = 0

( 3) hgrad = 0

( 4) wgrad = 0

F( 4, 36) = 10.06

Prob > F = 0.0000

Interaction Terms Applied:

( 1) agrad = 0

( 2) bgrad = 0

( 3) hgrad = 0

( 4) wgrad = 0

F( 4, 32) = 12.92

Prob > F = 0.0000

**Figure 6: F-Test (Interaction Terms Between Race and Income)**

( 1) c.lnincome#c.lnasian = 0

( 2) c.lnincome#c.lnblack = 0

( 3) c.lnincome#c.lnhisp = 0

( 4) c.lnincome#c.lnwhite = 0

F( 4, 32) = 1.67

Prob > F = 0.1807