# Testing Brueckner & Rosenthal: Did America's Downtowns Get Rich? *

Logan Cooper

April 30, 2023

## 1 Introduction

In 2008, Jan Brueckner and Stuart Rosenthal published a paper entitled "Gentrification and neighborhood housing cycles: will America's future downtowns be rich?" The central question of this paper is how people of different income levels sort themselves geographically in a city. Prior to this paper, most models framed this question as a trade-off between amenities and commute time into the Central Business District (CBD) in a monocentric city model (Brueckner et al., 1999; Alonso, 1964; Mills, 1967; Muth, 1969).

The major contribution of their paper was the addition of the age distribution of housing stock as a determinant of income in this model (Brueckner and Rosenthal, 2009). The authors find that with this addition, the positive effect of distance from the city center on income reverses, becoming negative instead of positive. In other words, given equally new housing in the suburbs and the CBD, the wealthy would rather live downtown.

Furthermore, the authors use this trend to extrapolate forward to 2020 and suggest that "[a]mong the largest cities, suburban income relative to the city center is expected to decline by roughly 10% of MSA mean income between 2000 and 2020, roughly one-fourth of the year 2000 disparity." (Brueckner and Rosenthal, 2009). Given the popular discourse about gentrification and housing scarcity in that time frame, it only makes sense to ask if this prediction came true.

Answering that question is the purpose of this paper. In it, I echo the methodology of Brueckner and Rosenthal's paper, but use the new data generated since 2008. Section 2 summarizes and justifies my data sources and empirical methodology. Section 3 summarizes my findings and compares them Brueckner and Rosenthal's predictions. Finally, section 4 concludes.

*Special thanks to Jenna Goldberg for help with the Census data.

## 2 Data & Methodology

Brueckner and Rosenthal used data from the 1980 and 2000 US Censuses sourced from the Neighborhood Change Database (NCDB). To test whether their predictions are accurate, I use data from the 2000-2019 period. I use a 19-year window instead of a 20-year window for two reasons. First, the NCDB does not go up to 2020, so any tract conversions from 2000 to 2020 would have needed to be done twice, once from 2000 to 2010, and then again from 2010 to 2020. Second, 2020 was an outlier of a year to say the least. Any changes in the income of people in the city center versus the suburbs were just as likely to be due to the COVID-19 pandemic as any underlying forces from housing age. This factor is compounded by the fact that in a year of almost exclusively remote work, balancing commute time against other factors was a non-issue for most highly-paid workers in urban areas.

To build the dataset, I begin with data on the share of workers commuting with public transit and on the distribution of housing ages from the 2000 NCDB (Tatian, 2003). In addition, I gathered the following data from the 2019 American Community Survey (ACS): median household income on both the MSA and tract levels, and population on the tract level. Also from the Census Bureau, I found a crosswalk for associating tracts with MSAs and another for converting 2000 tracts to 2010 tracts (2019) [1] . I also sourced a crosswalk for associating 2010 Census tracts with school districts from the National Center for Education Statistics (2019). Finally, for calculating the distance between tracts, I used the NBER Tract Distance Database (2010).

Identifying the center of a city is a somewhat contentious topic. Because of some unusual results from Brueckner and Rosenthal's method of identifying the CBD (discussed in section 3.1 and appendix A.1), I introduce a second method of doing so. Holian (2019) explores multiple methods of identifying a city center and finds that "the location of a city's city hall is a better proxy for the location of a city's central business district (CBD) than are other measures in current use". Therefore, I draw from Wilson (2012) who pioneered that method and gathered the data on city hall locations. This method uses the Census tract containing the city hall of the MSA's largest city as the CBD. We do this to remedy a weakness in the Brueckner and Rosenthal distance-based technique, which places CBDs in odd places for many cities – for example, placing Los Angeles' CBD in the city of Fullerton. To do this, I ran the latitude and longitude of each city hall – as calculated by Wilson – through the FCC Area API, which converts latitude and longitude measurements into Census FIPS codes.

In the rest of this paper, I will refer to the results of the Brueckner and Rosenthal methodology as BR-CBDs and BR-distance and the results of Wilson's method as Wilson-CBDs and Wilson-distance.

After joining the data (figure 1 illustrates the schema of these datasets – each arrow represents a data join), we have tract-level data for every (non-empty) 2010 census tract in every MSA in the United States within 40 miles of the CBD including the following information:

1. Tract GEOID for 2010.

[1] Because of changes to MSA codes between the 2000 and 2010 Censuses, some MSA codes had to be manually updated in the data to ensure that all of the data was present in the final dataset.

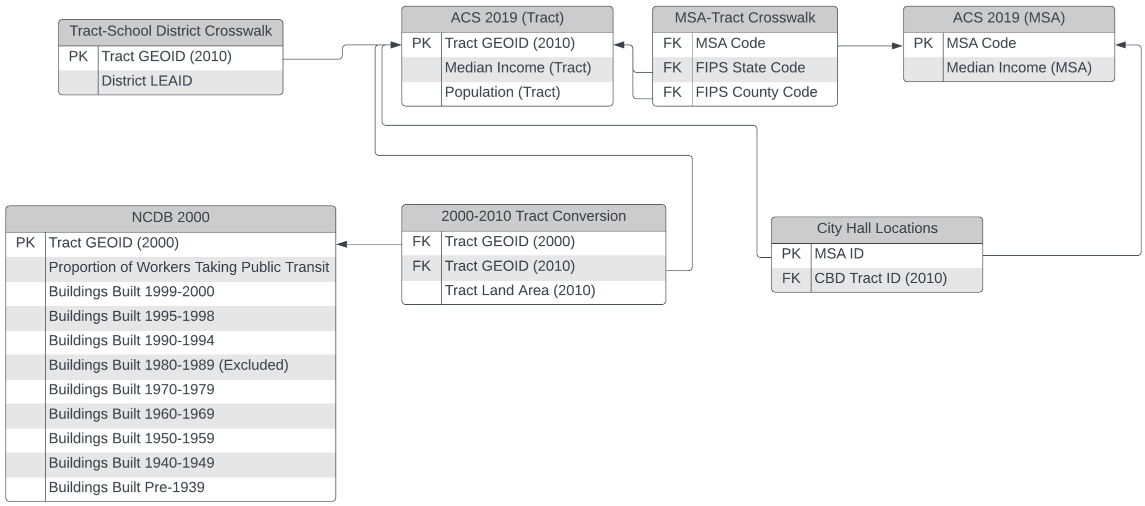



Figure 1: Dataset Schema

2. MSA ID for each tract – i.e. which MSA a tract belongs to.
3. Population density in 2019 (people per square mile); derived from tract population in 2019 and land area in 2010.
4. Median income of each tract relative to it's MSA in 2019; derived from median income in tract divided by median income in its MSA.
5. Whether 10% or more of a tract's workers took public transit to work in 2000; derived from the proportion of workers taking public transit in 2000.
6. A series of variables describing the share of buildings in a tract built in different time buckets (see figure 1 for the buckets) in 2000. Derived from number of buildings built in those buckets in 2000 [2].
7. The distance of a tract from its CBD; derived from the NBER tract distance database (absent from schema). CBD is defined two ways in the data: once using Brueckner and Rosenthal's method of choosing the tract with the highest population density, and another using Wilson's method of choosing the tract with the city hall.

I then run four weighted least squares (WLS) [3] linear regressions to replicate Brueckner and Rosenthal's results. Let $y_j$ denote a tract's ($j$) relative income to its MSA, $x_j$ its distance from the CBD, $\delta$ the coefficient on $x$, $\vec{m}$ a vector of MSA fixed effects, $\vec{\beta}$ the coefficient vector for $\vec{m}$, $t_j$ a dummy for ten percent or more of a tract's population taking public transit, $\zeta$ the

[2] First, 2000 tracts were converted to 2010 tracts. 2000 tracts which were created by splitting 2000 tracts all have the same values. Those which were created by combining tracts had their building counts summed up. Finally, the counts for all ages were summed up across each tract, and the age values were divided by the total to get percent values in the range $[0, 1]$.

[3] The reason for choosing WLS over ordinary least squares is the large degree of heteroskedasticity on the distance variable, see section 3 and specifically figure 2.

coefficient on $t_j$, $\vec{z_j}$ be a vector of school district fixed effects, $\vec{\eta}$ the vector of coefficients for $\vec{z_j}$, $\vec{a_j}$ the age distribution in 2000, $\vec{\theta}$ the coefficient for age, and finally $\omega_j$ the error term.

Following Brueckner and Rosenthal gives us four regression equations [4]. The first uses distance and MSA fixed effects to predict tract-income:

$$y_j = \delta_j x_j + \vec{\beta} \cdot \vec{m} + \omega_j$$

The second uses school-district fixed effects instead of MSA fixed effects:

$$y_j = \delta_j x_j + \vec{\eta} \cdot \vec{z_j} + \omega_j$$

The third adds the public transit dummy variable to equation two:

$$y_j = \delta_j x_j + \zeta_j t_j + \vec{\eta} \cdot \vec{z_j} + \omega_j$$

Finally, the fourth adds the age variables to equation three [5]:

$$y_j = \delta_j x_j + \zeta_j t_j + \vec{\eta} \cdot \vec{z_j} + \vec{\theta} \cdot \vec{a_j} + \omega_j$$

Because we use two different measures of distance from two different definitions of the CBD, each of these regressions must be run twice: once with BR-distances and once with Wilson-distances.

Additionally, I also add fifth and sixth regressions which add square and cubic terms respectively to equation four. This is to handle non-linearity present in the distance relationships. The additional coefficients and variables should be self-evident, and the final (i.e. sixth) regression equation with both terms is written below:

$$y_j = \delta_{j1} x_j + \delta_{j2} x_j^2 + \delta_{j3} x_j^2 + \zeta_j t_j + \vec{\eta} \cdot \vec{z_j} + \vec{\theta} \cdot \vec{a_j} + \omega_j$$

Additionally, we split the data into four tranches based on the number of census tracts in an MSA. This gives us four subsets: one with MSAs with less than 100 tracts, one with MSAs with 100-499 tracts, one with MSAs with 500-999 tracts, and finally one with 1000 or more tracts. Therefore, with six equations, two metrics, and five subsets (including the whole dataset), we are left with a total of 60 regression models.

[4]For computational reasons, the actual order of the variables in the code may differ from the notation here. This order is primarily to match Brueckner and Rosenthal's regression equations and to highlight the differences between equations.

[5]As noted in figure 1, I omit the variable for the share of buildings built between 1980 and 1989 from $\vec{a}$. This follows Brueckner and Rosenthal's methodology in which they remove buildings which were between 30 and 39 years old from their sample.

# 3 Findings

## 3.1 Exploratory Analysis

From just a cursory analysis, one can observe a major difference between the BR and Wilson distances. This expresses itself in differences in CBD-location, distribution, and even sample size. We can see from the descriptive statistics in table 1 that although the income variables are similar, the BR-distance has a higher mean and standard deviation than the Wilson-distances.

From graphing these distributions, we can also see that the BR distribution is relatively flat (figure 2a), with large clusters around the MSA median at the center and between 10 and 20 miles away (see 2c). The Wilson distances meanwhile display noted heteroskedasticity (figure 2b), and tend to cluster around the MSA median, while both tapering off in income and petering out in density as the distance increases (figure 2d).

Table 1: Descriptive Statistics for Both Distance Measures

(a) Descriptive Statistics for BR Distance

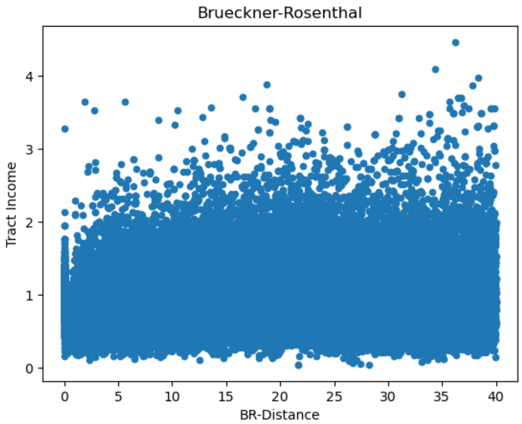
| | Income | Distance |
|---|---|---|
| Count | 56028 | 56028 |
| Mean | 1.072 | 20.285 |
| Std | 0.434 | 10.552 |
| Min | 0.041 | 0.000 |
| 25% | 0.774 | 11.861 |
| 50% | 1.022 | 20.045 |
| 75% | 1.300 | 28.857 |
| Max | 4.463 | 39.999 |

(b) Descriptive Statistics for Wilson Distance

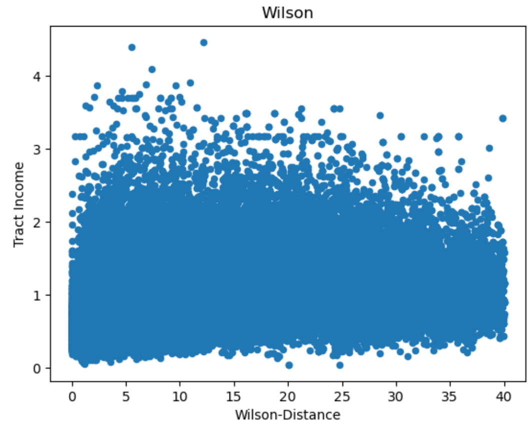
| | Income | Distance |
|---|---|---|
| Count | 77055 | 77055 |
| Mean | 1.078 | 13.729 |
| Std | 0.447 | 9.651 |
| Min | 0.041 | 0.000 |
| 25% | 0.769 | 5.840 |
| 50% | 1.023 | 11.557 |
| 75% | 1.310 | 20.005 |
| Max | 4.463 | 39.998 |

The BR CBDs also tend not to be very central, as I mentioned in section 2. Relying on population density leads to some oddly-placed city centers. A sample of these from Southern California can be seen in figure A1. This also holds they key to why the sample size is so much lower for BR-distance. I limit the distance a tract can be from the city center to 40 miles to emulate Brueckner and Rosenthal's approach. When CBDs are not central, large parts of an MSA can be excluded with that criterion. Take for example Bakersfield, which is not in the sample (note its lack of color in figure A1.) That is because its BR-CBD is a good distance to the North, off the map, such that the main city itself is too far away to be in the sample. Because of this problem, any results from BR distances should be approached with caution.

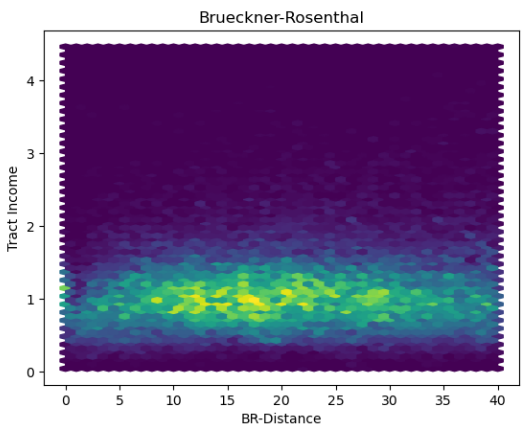
Using Wilson distances gets around this problem nicely. The map coverage is much greater, even if it loses some tracts on the edges of certain areas (note Orange County in figure A2). This improved centrality also leads to more sensible estimation results, as we will see in the next subsection.



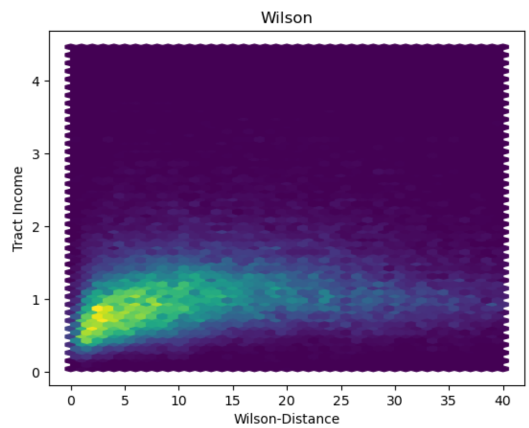

(a) Scatter Plot for BR-Distance




(b) Scatter Plot for Wilson-Distance




(c) Hex Plot for BR-Distance




(d) Hex Plot for Wilson-Distance

Figure 2: Graphical Conditional Distributions of Income on Distance

## 3.2 Regression Results

The central result of all of my regressions is that the effect that Brueckner and Rosenthal describe – the sign of the coefficient of distance on income flipping from positive to negative upon the addition of housing age – fails to hold up when using modern data and their density-based criterion of centrality. Their results hold up better – but still not perfectly – when using city hall location.

### 3.2.1 BR-Distance Regressions

For MSAs with under 100 tracts, regression one gives us a highly significant ($p < 0.001$) coefficient of 0.0040. This goes up as we add amenity controls (school district fixed effects and public transit use), and then down slightly to 0.0031 when adding the age distribution (the coefficients of which are spread between either significant at below the 1% level or insignificant at the 5% level). Adding non-linear terms greatly increases the coefficient on distance, but as we will see in subsection, 3.3, the non-linear terms are likely misspecified for both the under-100 and 100-499 subsets.

The 100-499 subset shows an entirely different behavior. Once again, almost all of the coefficients (with the exception of distance when including distance-cubed and a few of the age coefficients) are significant at lower than the 0.1% level. However, the distance coefficient starts out relatively weak and gets stronger as more features are added. The 500-999 subset start negative (-0.0019, $p = 0.004$), and gets less negative as amenities are added, but then goes back to near its original value as dwelling age is added. The addition of non-linear terms then turns the distance coefficient positive. Note that the distance coefficients for everything but regressions one and six are insignificant even at the 10% level.

Finally, for the largest cities with over 1000 tracts, we see distance start positive and get stronger with the addition of more variables, until the addition of distance-cubed switches the sign. All of the distance coefficients are extremely significant.

#### 3.2.2 Wilson-Distance Regressions

Using Wilson-distance leads to much more consistent and sensible results. When looking at all MSAs in the sample, we see the distance coefficient start out relatively strong at 0.0097, and drop steadily as amenities (0.0092), transit (0.007), and age (0.0024) are added. The addition of distance-squared actually increases the coefficient of age by an order of magnitude – to 0.0243, but with a negative squared term, which matches Brueckner and Rosenthal's results: a negative parabolic relationship between distance and income. All of the coefficients are highly significant.

In the smallest cities (under 100 tracts), we see the distance-income relationship start off stronger at 0.0062, and strengthen with the addition of the amenity measures before dropping precipitously upon adding dwelling age. Adding distance-squared, we see the same strengthening of the distance but with a (very weakly) negative distance-squared coefficient, matching the negative-parabolic effect. The distance-cubed regression is likely misspecified.

In 100-499 tract and 500-999 tract MSAs, we see a very similar pattern. Distance's coefficient increases as amenities are added, decreases with dwelling age, and gets very strongly positive but negatively parabolic with distance-squared. Cities with 1000 or more tracts most closely match the pattern noted in Brueckner and Rosenthal, with the distance coefficient starting off strong and getting monotonically weaker with more features until it finally turns negative with distance-cubed.

Table 2: Regression Results: All MSAs for BR-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.004 (<0.001) | 0.0058 (<0.001) | 0.0053 (<0.001) | 0.004 (<0.001) | 0.0081 (<0.001) | 0.0082 (<0.001) |
| Distance-Squared | - | - | - | - | -0.0001 (<0.001) | -0.0001 (0.309) |
| Distance-Cubed | - | - | - | - | - | 1.103e-07 (0.951) |
| Access to Public Transit in 2000 | - | - | -0.3181 (<0.001) | -0.2714 (<0.001) | -0.2714 (<0.001) | -0.2714 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.9036 (<0.001) | 0.9037 (<0.001) | 0.9037 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.4814 (<0.001) | 0.481 (<0.001) | 0.481 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.7542 (<0.001) | 0.7529 (<0.001) | 0.7529 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.1637 (<0.001) | -0.1631 (<0.001) | -0.1632 (<0.001) |
| % Buildings Built 1960-1969 | - | - | - | -0.2218 (<0.001) | -0.2186 (<0.001) | -0.2186 (<0.001) |
| % Buildings Built 1950-1959 | - | - | - | 0.0771 (0.002) | 0.082 (0.001) | 0.082 (0.001) |
| % Buildings Built 1940-1949 | - | - | - | -0.6738 (<0.001) | -0.6738 (<0.001) | -0.6738 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | -0.0771 (0.001) | -0.0748 (0.001) | -0.0748 (0.001) |
| Intercept | 1.0318 (<0.001) | 0.9792 (<0.001) | 0.9813 (<0.001) | 1.0215 (<0.001) | 1.0043 (<0.001) | 1.004 (<0.001) |
| Observations | 56028 | 56028 | 56028 | 56028 | 56028 | 56028 |
| MSA Fixed Effects | 349 | - | - | - | - | - |
| School District Fixed Effects | - | 5484 | 5484 | 5484 | 5484 | 5484 |
| Adjusted $R^2$ | 0.029 | 0.296 | 0.322 | 0.394 | 0.394 | 0.394 |

Table 3: Regression Results: All MSAs for Wilson-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.0097 (<0.001) | 0.0092 (<0.001) | 0.007 (<0.001) | 0.0024 (<0.001) | 0.0243 (<0.001) | 0.047 (<0.001) |
| Distance-Squared | - | - | - | - | -0.0007 (<0.001) | -0.0023 (<0.001) |
| Distance-Cubed | - | - | - | - | - | 2.922e-05 (<0.001) |
| Access to Public Transit in 2000 | - | - | -0.2591 (<0.001) | -0.2275 (<0.001) | -0.2103 (<0.001) | -0.2036 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.8482 (<0.001) | 0.8148 (<0.001) | 0.8209 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.5147 (<0.001) | 0.4889 (<0.001) | 0.4855 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.7207 (<0.001) | 0.7094 (<0.001) | 0.7046 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.1479 (<0.001) | -0.149 (<0.001) | -0.1495 (<0.001) |
| % Buildings Built 1960-1969 | - | - | - | -0.2375 (<0.001) | -0.2103 (<0.001) | -0.2042 (<0.001) |
| % Buildings Built 1950-1959 | - | - | - | 0.1162 (<0.001) | 0.1347 (<0.001) | 0.141 (<0.001) |
| % Buildings Built 1940-1949 | - | - | - | -0.7427 (<0.001) | -0.7051 (<0.001) | -0.6979 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | 0.0476 (0.013) | 0.1259 (<0.001) | 0.1709 (<0.001) |
| Intercept | 0.9718 (<0.001) | 0.9667 (<0.001) | 0.9759 (<0.001) | 1.0173 (<0.001) | 0.9252 (<0.001) | 0.8562 (<0.001) |
| Observations | 77055 | 77055 | 77055 | 77055 | 77055 | 77055 |
| MSA Fixed Effects | 349 | - | - | - | - | - |
| School District Fixed Effects | - | 6850 | 6850 | 6850 | 6850 | 6850 |
| Adjusted $R^2$ | 0.045 | 0.305 | 0.322 | 0.389 | 0.396 | 0.398 |

### 3.3 Did City Centers Get Rich?

While the nature of the relationship between wealth, distance, amenities, and housing age is of interest in and of itself, the most provocative claim of Brueckner and Rosenthal's is this: "Although our model forecasts a reduction in the central-city/suburb disparity in economic status, central cities will nevertheless continue to be poorer than suburban communities" (2009). Later on, they also note that this effect is likely to be strongest in the largest cities. In light of popular discourse about housing prices and gentrification in many cities across the United States, such predictions seem very prescient.

Brueckner and Rosenthal made this prediction by using their model to predict 2020's cities from 2000's data, and illustrated the point with a specific set of figures, showing the mean income relative to the CBD at different distances. These predictions show notable non-linearity, hence the addition of square and cubic terms in this paper. To evaluate their claims, I replicate these charts with modern data by binning the distances into discrete (mostly) one-mile intervals by rounding the distance numbers to the nearest whole number [6]. The remainder of this section will be dedicated to a discussion of these figures.

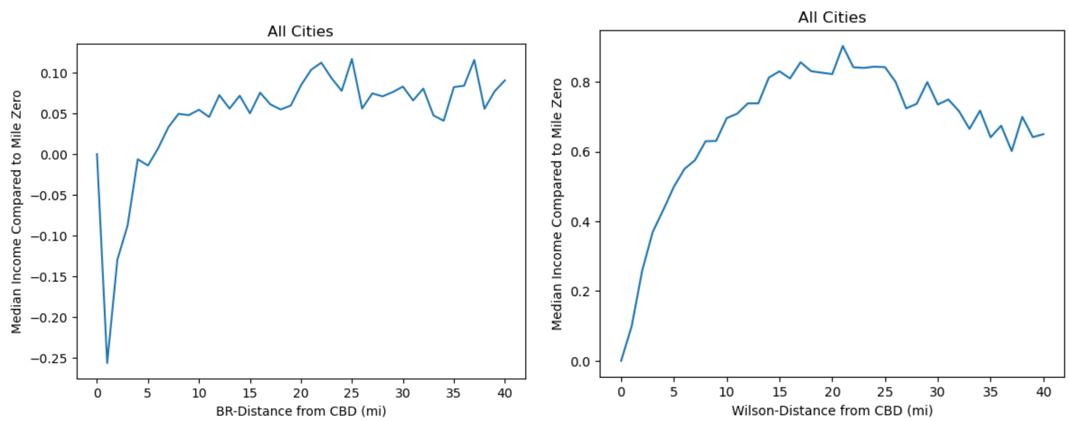



Figure 3: Income and Binned-Distance for All MSAs

From figure 3 above, a problem with using the BR distances becomes immediately apparent. Because the BD-CBDs are often far outside the 'actual' city center, we see the income plummet immediately outside of the center, then recover and stay relatively constant slightly above the level of the CBD. This problem persists with slight differences in severity in all of the subsets. Wilson distances however show us a picture where incomes rise sharply out of the CBD, peak at around 20-25 miles out, and then begin to decline. This maps nicely to the understanding of US cities having very poor city centers, rich suburbs, then a slow transition to poorer exurban and rural areas. While on the surface the large and regular discrepancy between central city incomes and suburban ones could be read as a refutation of Brueckner and Rosenthal, variation appears in this relationship when the sample is broken down by MSA size.

For small cities (see figure 4) the BR-distance figure resembles that for all cities very closely,

[6] I also tested with with floor and ceiling functions, and the difference between the methods was minimal.

albeit with more volatility. For Wilson-distance, we see the income shoot up much more quickly and to a higher point, peaking at around twice the median income of the CBD at 10 miles, then descend much more quickly to a lower level than we saw in the overall sample. This can be explained by very small MSAs being – almost by definition – closer to rural areas than larger cities. In this case, the relationship between distance and income for small cities is the same that Brueckner and Rosenthal describe, but it has gotten even more extreme: their paper predicts a similar, but far weaker, pattern: one that peaks at just over 0.3 at around 7 miles before descending back to near the city center value.

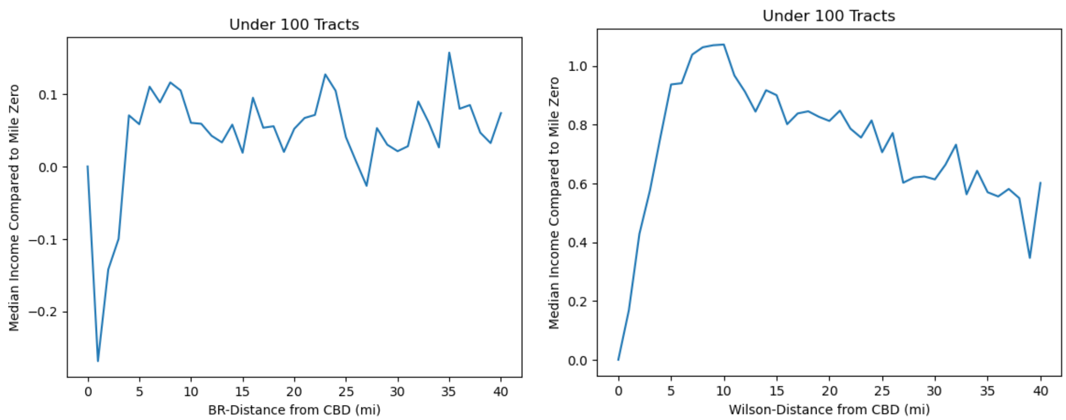



Figure 4: Income and Binned-Distance for Small (< 100 Tract) MSAs

For MSAs with 100-499 tracts measured with the Wilson distance, the relationship is nearly the same as for small MSAs, but the whole distribution is shifted right, such that the come-up is slower and more tapered and the peak is just after 10 miles out.

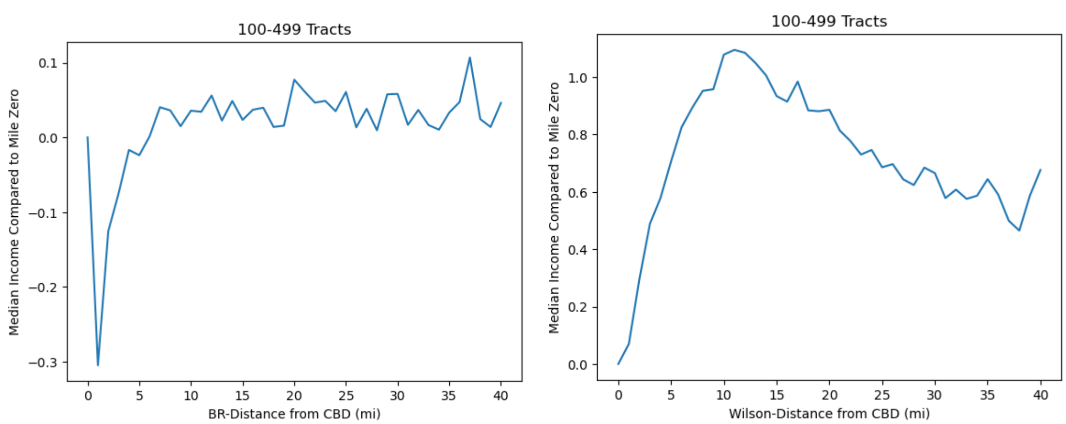



Figure 5: Income and Binned-Distance for Medium (100-499 Tract) MSAs

It is not until we get to large MSAs that we see variation in this pattern (see figure 6). Surprisingly, we see the same initial drop-off for both measures, akbeit weaker and less steep in the Wilson case. This is consistent with the concentric model of city growth posited by Brueckner and Rosenthal. If city centers are the first places to be renovated and gentrified (here simply meaning rich residents moving in) then we should see a drop-off like this between the newly-rich city center and the first ring of housing outside of it which has not been renovated and is still home to its original residents.

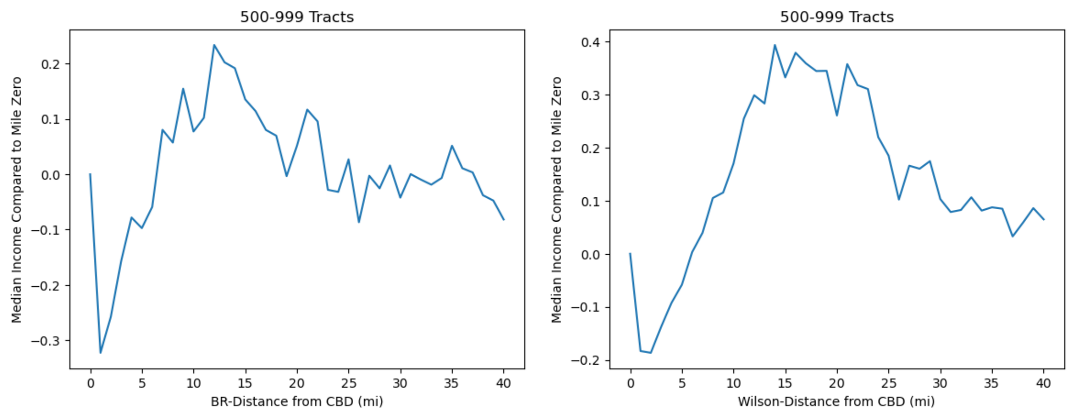



Figure 6: Income and Binned-Distance for Large (500-999 Tract) MSAs

Although we see this drop-off with both measures, the sharpness of it in the BR-measure and the fact that it is always present are good reasons to still be skeptical of this measure. While it is not exact, this does match the predicted behavior for these large MSAs. However, it more closely matches the predicted behavior for very large ($\geq$ 1000 Tracts) MSAs, where Brueckner and Rosenthal predict a similar sort of dip (see figure 7).

Very large cities are the case in which BR- and Wilson- distances match up best, which is surprising considering where the BR method places the center in the mapped cases of Los Angeles and San Diego, as well as the un-mapped San Jose and San Francisco (see appendix A.1). For both measures, we see income drop (despite an initial jump in the Wilson chart) just outside of the center before rising again to levels between 0.3 and 0.4, which matches Brueckner and Rosenthal's predictions nicely. When measuring with the BR-distance, we see income rising sharply at the very edge of the chart, after around 37 miles. Because of the non-central CBDs generated by this method, we likely see this due to rich city centers being that far from the BR city center in some samples (e.g. Los Angeles.)

The Wilson measure matches Brueckner and Rosenthal's predictions almost exactly, down to the small increase in income around 1-2 miles out. I hypothesize that this is due to the placement of city halls in large cities. Most are not in areas with residential construction: see New York City Hall in the City Center area surrounded by parks and civic buildings, Los Angeles City Hall in Downtown LA surrounded by office buildings and a large library, or Chicago City Hall.

Overall, Brueckner and Rosenthal's predictions hold true for larger cities, but not smaller ones. In fact, the disparity between the wealth of city centers and suburbs seems to have increased in cities with under 500 tracts. I speculate that this is due to a sort of geographic substitution effect. Due to chronic under-building of housing since the 2008 Financial Crisis (Chernick et al., 2011), many Americans have moved from high to low cost of living areas. Take for

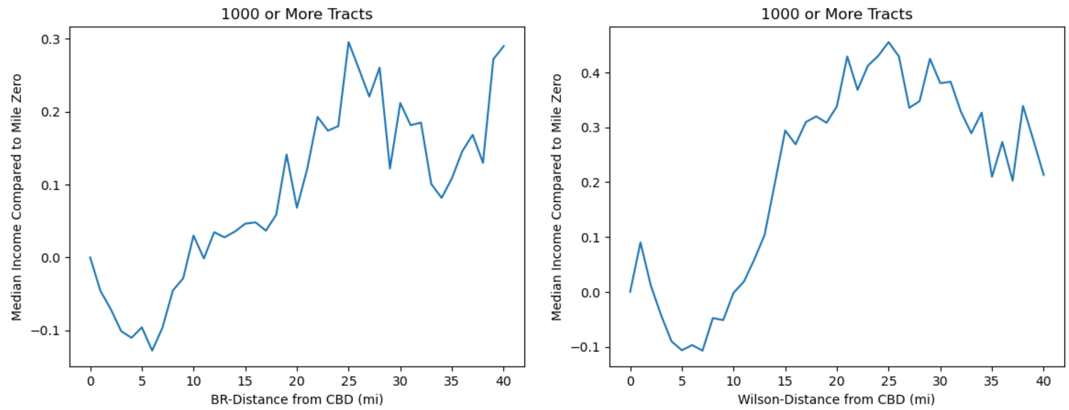



Figure 7: Income and Binned-Distance for Very Large ($\geq$ 1000 Tracts) MSAs

example California's slowing population growth and recent population decline (Hans Johnson and Mejia, 2023) and the popular conception of people moving from California to mid-sized cities in places like Idaho, Texas, and Arizona (Brand et al., 2021; Barringer, 2022). If this trend is a general one, and people who are relatively poor in their home city are moving to places where they are relatively rich – specifically suburbs of those areas – this would lead (all else equal) to higher income discrepancies in those smaller cities, especially if those moving are highly-skilled relative to the current residents or if they are able to work remotely for salaries not easily attainable in their destination cities.

To take this one step further, I would also speculate that people moving from large cities to small or mid-sized ones would move to the suburbs, as I would posit that small cities are more likely than large ones to have car-dependent, low-amenity downtowns. Therefore, an influx of wealthier people to small-town suburbs chould create the pattern we see in figures 4 and 5.

# 4 Conclusion

Overall, Brueckner and Rosenthal's predictions hold true for the larger cities in the sample, but not the smaller ones. Additionally, the sign-flipping result for the effect of distance seems to have disappeared. This may be due to non-linear relationships in the new data that were not present (or at least less prominent) in the year 2000. I attempted to capture this by including square and cubic terms on distance, and while the squared regressions seem to work well, the cubic ones are likely misspecified. Future research might try additional ways of capturing this non-linearity, for example log-models or local regression.

Additionally, data problems may play a role in these results. Brueckner and Rosenthal were able to rely on the Neighborhood Change Database for their whole time period, whereas I was forced to generate approximations of that data directly from Census data. A more granular approach to doing so (i.e. using changes in blocks to build up changes in tracts) like the NCDB does might lead to clearer results. In a similar vein, but one more critical of the original research design, using unified school districts may not be a good measure of local amenities for many large cities. New York and Los Angeles Proper are served by one school district each, with huge disparities between their best and worst schools/neighborhoods. A better measure might be one which breaks down districts into individual school zones (e.g. which tracts are served by which high schools). Such a research design would be highly computationally intensive (likely with tens of thousands of school fixed effects) but would capture far more intra-city variation.

In addition to fixing those possible problems with my research design (or assuming my results are valid), digging into the differences in behavior between MSAs under 500 tracts and those over 500 tracts could be a fruitful direction for future research. Specifically, testing my hypothesis set out in the last two paragraphs of section 3.3 would be an interesting avenue.

The title of this paper put forward a question: Did American's downtowns get rich? Given the current data and methodology, the best short answer seems to be: "It depends." For large cities: yes; for smaller ones: no [7].

[7] That's it! Grad school is done. Wrap it up folks. We're outta here.

## References


Alonso, W. (1964). *Location and land use: toward a general theory of land rent*. Harvard university press.

Barringer, F. (2022). Idaho is increasingly a bullseye for californian migrants. *the West*.

Brand, J., N. Fultz, B. Freyd, S. Gabriel, and M. Tzen (2021). Recent trends in california migration evidence from the american community survey 2005-2019.

Brueckner, J. K. and S. S. Rosenthal (2009). Gentrification and neighborhood housing cycles: will america's future downtowns be rich? *The Review of Economics and Statistics 91*(4), 725–743.

Brueckner, J. K., J.-F. Thisse, and Y. Zenou (1999). Why is central paris rich and downtown detroit poor?: An amenity-based theory. *European economic review 43*(1), 91–107.

Chernick, H., A. Langley, and A. Reschovsky (2011). The impact of the great recession and the housing crisis on the financing of america's largest cities. *Regional Science and Urban Economics 41*(4), 372–381.

Hans Johnson, E. M. and M. C. Mejia (2023). Fact sheet: California's population.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant (2020, September). Array programming with NumPy. *Nature 585*(7825), 357–362.

Holian, M. J. (2019). Where is the city's center? five measures of central location. *Cityscape 21*(2), 213–226.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering 9*(3), 90–95.

Jordahl, K., J. V. den Bossche, M. Fleischmann, J. Wasserman, J. McBride, J. Gerard, J. Tratner, M. Perry, A. G. Badaracco, C. Farmer, G. A. Hjelle, A. D. Snow, M. Cochran, S. Gillies, L. Culbertson, M. Bartos, N. Eubank, maxalbert, A. Bilogur, S. Rey, C. Ren, D. Arribas-Bel, L. Wasser, L. J. Wolf, M. Journois, J. Wilson, A. Greenhall, C. Holdgraf, Filipe, and F. Leblanc (2020, July). geopandas/geopandas: v0.8.1.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review 57*(2), 197–210.

Muth, R. F. (1969). Cities and housing; the spatial pattern of urban residential land use.

NBER (2010). Tract distance database. https://www.nber.org/research/data/tract-distance-database.

NCES (2019). School district geographic relationship files. https://nces.ed.gov/programs/edge/Geographic/RelationshipFiles.

pandas development team, T. (2020, February). pandas-dev/pandas: Pandas.

Seabold, S. and J. Perktold (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Tatian, P. A. (2003, October). Censuscd neighborhood change database (ncdb) 1970 – 2000 tract data data users' guide long form release. Technical report, The Urban Institute in collaboration with GeoLytics Inc., The Urban Institute 2100 M Street, NW Washington, DC 20037.

USCB (2019). American community survey, 2019.

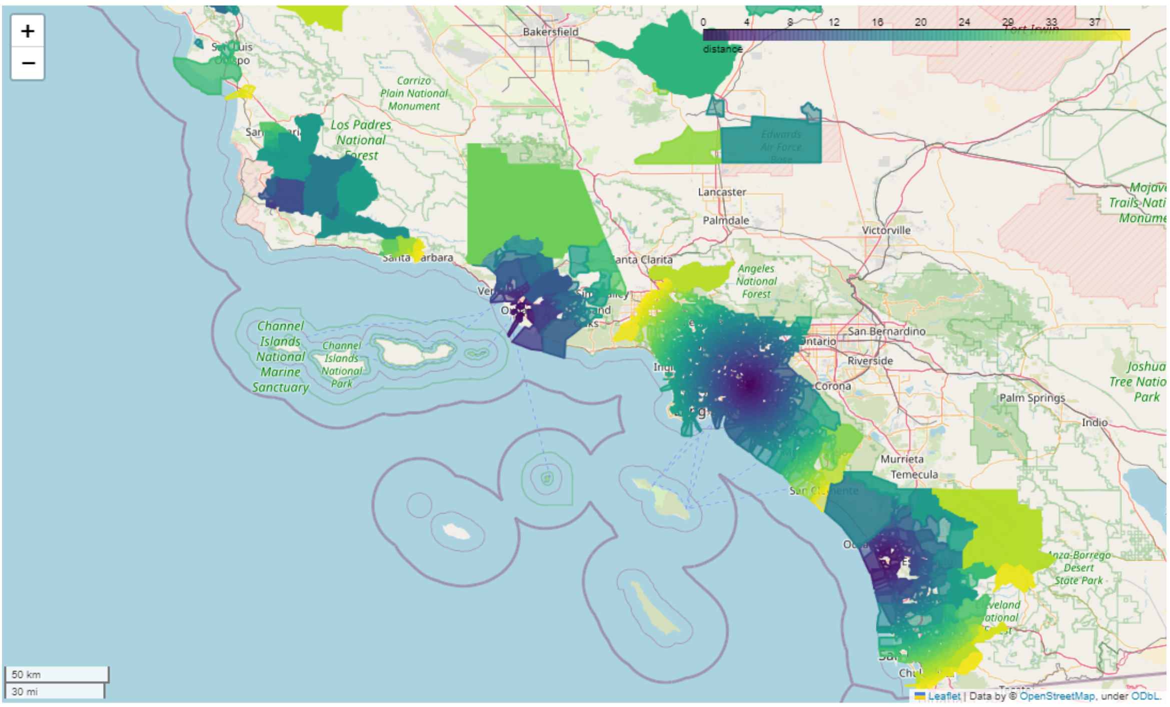
Wes McKinney (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference*, pp. 56 – 61.

Wilson, S. G. (2012). *Patterns of metropolitan and micropolitan population change: 2000 to 2010.* US Department of Commerce, Economics and Statistics Administration, US ....

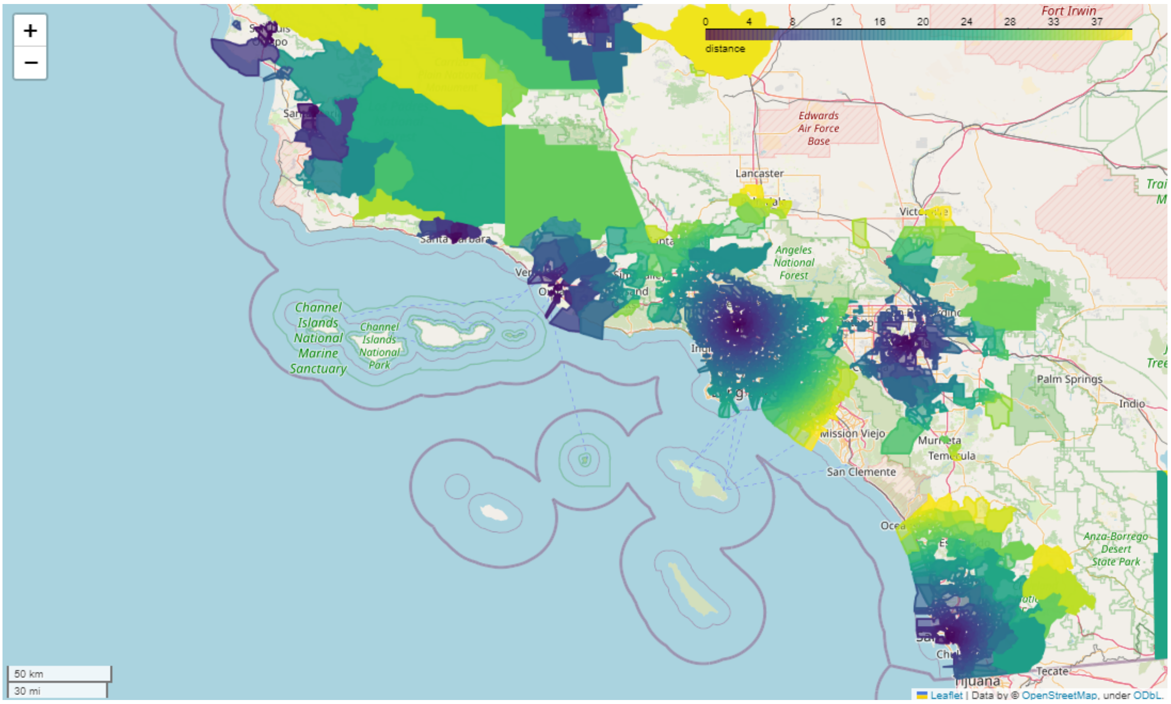
# A Appendices

All of the code for this project can be found here on GitHub, which functions as a sort of online appendix. All of the programming for this paper was done in Python 3.11.2 with the help of the libraries pandas (pandas development team, 2020; Wes McKinney, 2010), numpy (Harris et al., 2020), statsmodels (Seabold and Perktold, 2010), matplotlib (Hunter, 2007), and geopandas (Jordahl et al., 2020).

## A.1 Maps




Appendix Figure A1: A map of BR-distance by census tract in Southern California




Appendix Figure A2: A map of Wilson-distance by census tract in Southern California

## A.2 Subset Regression Tables

Appendix Table A1: Regression Results: MSAs Under 100 Tracts for BR-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.004 (<0.001) | 0.0043 (<0.001) | 0.0041 (<0.001) | 0.0031 (<0.001) | 0.008 (<0.001) | 0.0181 (<0.001) |
| Distance-Squared | - | - | - | - | -0.0001 (<0.001) | -0.0009 (<0.001) |
| Distance-Cubed | - | - | - | - | - | 1.335e-05 (<0.001) |
| Access to Public Transit in 2000 | - | - | -0.2822 (<0.001) | -0.1708 (<0.001) | -0.1705 (<0.001) | -0.1704 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.9865 (<0.001) | 0.9956 (<0.001) | 0.9876 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.2933 (0.001) | 0.2874 (0.001) | 0.288 (0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.9934 (<0.001) | 0.9927 (<0.001) | 0.9816 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.2088 (0.008) | -0.2108 (0.007) | -0.2189 (0.005) |
| % Buildings Built 1960-1969 | - | - | - | -0.3164 (<0.001) | -0.3148 (<0.001) | -0.3156 (<0.001) |
| % Buildings Built 1950-1959 | - | - | - | -0.1025 (0.147) | -0.0971 (0.169) | -0.1007 (0.154) |
| % Buildings Built 1940-1949 | - | - | - | -1.0343 (<0.001) | -1.0282 (<0.001) | -1.0318 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | -0.4379 (<0.001) | -0.4348 (<0.001) | -0.4308 (<0.001) |
| Intercept | 1.0303 (<0.001) | 0.9471 (<0.001) | 0.9497 (<0.001) | 0.971 (<0.001) | 0.931 (<0.001) | 0.8989 (<0.001) |
| Observations | 11325 | 11325 | 11325 | 11325 | 11325 | 11325 |
| MSA Fixed Effects | 208 | - | - | - | - | - |
| School District Fixed Effects | - | 1822 | 1822 | 1822 | 1822 | 1822 |
| Adjusted $R^2$ | 0.024 | 0.103 | 0.115 | 0.295 | 0.296 | 0.297 |

Appendix Table A2: Regression Results: MSAs 100-499 Tracts for BR-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.0027 (<0.001) | 0.0059 (<0.001) | 0.0062 (<0.001) | 0.0046 (<0.001) | 0.0058 (<0.001) | 0.0027 (0.353) |
| Distance-Squared | - | - | - | - | 3.29e-05 (<0.001) | 0.0002 (0.307) |
| Distance-Cubed | - | - | - | - | - | 2.62e-06 (<0.001) |
| Access to Public Transit in 2000 | - | - | -0.3821 (<0.001) | -0.2914 (<0.001) | -0.2911 (<0.001) | -0.2908 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.8561 (<0.001) | 0.855 (<0.001) | 0.8523 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.3933 (<0.001) | 0.3934 (<0.001) | 0.3938 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.7359 (<0.001) | 0.7353 (<0.001) | 0.736 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.1743 (<0.001) | -0.1741 (<0.001) | -0.1736 (<0.001) |
| % Buildings Built 1960-1969 | - | - | - | -0.3104 (<0.001) | -0.3088 (<0.001) | -0.3087 (<0.001) |
| % Buildings Built 1950-1959 | - | - | - | 0.0775 (0.063) | 0.0782 (0.060) | 0.0783 (0.060) |
| % Buildings Built 1940-1949 | - | - | - | -0.7615 (<0.001) | -0.7608 (<0.001) | -0.7612 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | -0.1472 (<0.001) | -0.1461 (<0.001) | -0.1461 (<0.001) |
| Intercept | 1.0431 (<0.001) | 0.9787 (<0.001) | 0.9775 (<0.001) | 1.0617 (<0.001) | 1.0564 (<0.001) | 1.0664 (<0.001) |
| Observations | 24584 | 24584 | 24584 | 24584 | 24584 | 24584 |
| MSA Fixed Effects | 120 | - | - | - | - | - |
| School District Fixed Effects | - | 2526 | 2526 | 2526 | 2526 | 2526 |
| Adjusted $R^2$ | 0.042 | 0.254 | 0.288 | 0.378 | 0.378 | 0.378 |

Appendix Table A3: Regression Results: MSAs 500-999 Tracts for BR-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | -0.0019 (0.004) | -0.0006 (0.649) | -0.0002 (0.852) | -0.0018 (0.147) | 0.0033 (0.343) | 0.0197 (0.006) |
| Distance-Squared | - | - | - | - | -0.0001 (0.117) | -0.0011 (0.004) |
| Distance-Cubed | - | - | - | - | - | 1.529e-05 (0.009) |
| Access to Public Transit in 2000 | - | - | -0.323 (<0.001) | -0.3085 (<0.001) | -0.3074 (<0.001) | -0.3074 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.5786 (<0.001) | 0.5816 (<0.001) | 0.5837 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.6182 (<0.001) | 0.6239 (<0.001) | 0.6164 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.6028 (<0.001) | 0.6067 (<0.001) | 0.5978 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.3594 (<0.001) | -0.3556 (<0.001) | -0.3614 (<0.001) |
| % Buildings Built 1960-1969 | - | - | - | -0.1429 (0.015) | -0.1371 (0.020) | -0.1419 (0.016) |
| % Buildings Built 1950-1959 | - | - | - | 0.059 (0.350) | 0.0656 (0.300) | 0.066 (0.297) |
| % Buildings Built 1940-1949 | - | - | - | -0.6673 (<0.001) | -0.6612 (<0.001) | -0.6617 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | 0.1899 (0.001) | 0.198 (<0.001) | 0.2069 (<0.001) |
| Intercept | 1.051 (<0.001) | 1.1345 (<0.001) | 1.1357 (<0.001) | 1.1228 (<0.001) | 1.0692 (<0.001) | 1.0109 (<0.001) |
| Observations | 7546 | 7546 | 7546 | 7546 | 7546 | 7546 |
| MSA Fixed Effects | 11 | - | - | - | - | - |
| School District Fixed Effects | - | 504 | 504 | 504 | 504 | 504 |
| Adjusted $R^2$ | 0.018 | 0.337 | 0.364 | 0.427 | 0.427 | 0.428 |

Appendix Table A4: Regression Results: MSAs 1000 or More Tracts for BR-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.0093 (<0.001) | 0.0191 (<0.001) | 0.0149 (<0.001) | 0.0137 (<0.001) | 0.0263 (<0.001) | -0.0214 (0.002) |
| Distance-Squared | - | - | - | - | -0.0003 (0.001) | 0.0027 (<0.001) |
| Distance-Cubed | - | - | - | - | - | 6.06e-06 (<0.001) |
| Access to Public Transit in 2000 | - | - | -0.3571 (<0.001) | -0.364 (<0.001) | -0.3645 (<0.001) | -0.3625 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 1.283 (<0.001) | 1.3016 (<0.001) | 1.321 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.6483 (<0.001) | 0.6402 (<0.001) | 0.646 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.6938 (<0.001) | 0.6896 (<0.001) | 0.703 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.0302 (0.634) | -0.0324 (0.609) | -0.0199 (0.753) |
| % Buildings Built 1960-1969 | - | - | - | -0.0286 (0.614) | -0.0242 (0.669) | -0.0159 (0.778) |
| % Buildings Built 1950-1959 | - | - | - | 0.2855 (<0.001) | 0.2965 (<0.001) | 0.3055 (<0.001) |
| % Buildings Built 1940-1949 | - | - | - | -0.2786 (<0.001) | -0.2895 (<0.001) | -0.2624 (0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | 0.2759 (<0.001) | 0.2642 (<0.001) | 0.2674 (<0.001) |
| Intercept | 1.1333 (<0.001) | -0.1434 (0.274) | 0.212 (0.100) | 0.13 (0.332) | 0.0349 (0.798) | 0.1846 (0.179) |
| Observations | 12573 | 12573 | 12573 | 12573 | 12573 | 12573 |
| MSA Fixed Effects | 7 | - | - | - | - | - |
| School District Fixed Effects | - | 792 | 792 | 792 | 792 | 792 |
| Adjusted $R^2$ | 0.162 | 0.466 | 0.491 | 0.513 | 0.514 | 0.516 |

Appendix Table A5: Regression Results: MSAs Under 100 Tracts for Wilson-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.0062 (<0.001) | 0.0086 (<0.001) | 0.0081 (<0.001) | 0.0039 (<0.001) | 0.0339 (<0.001) | 0.0764 (<0.001) |
| Distance-Squared | - | - | - | - | -0.001 (<0.001) | -0.0043 (<0.001) |
| Distance-Cubed | - | - | - | - | - | 6.49e-05 (<0.001) |
| Access to Public Transit in 2000 | - | - | -0.262 (<0.001) | -0.1524 (<0.001) | -0.1179 (<0.001) | -0.0984 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 1.0388 (<0.001) | 0.9097 (<0.001) | 0.8423 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.1908 (0.028) | 0.1007 (0.236) | 0.0956 (0.254) |
| % Buildings Built 1990-1994 | - | - | - | 1.0254 (<0.001) | 0.9887 (<0.001) | 0.9778 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.2349 (0.002) | -0.2596 (0.001) | -0.2368 (0.001) |
| % Buildings Built 1960-1969 | - | - | - | -0.2723 (<0.001) | -0.1938 (0.004) | -0.1233 (0.062) |
| % Buildings Built 1950-1959 | - | - | - | -0.1447 (0.039) | -0.0855 (0.213) | -0.0299 (0.659) |
| % Buildings Built 1940-1949 | - | - | - | -1.0046 (<0.001) | -0.8573 (<0.001) | -0.7232 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | -0.4532 (<0.001) | -0.3977 (<0.001) | -0.268 (<0.001) |
| Intercept | 1.0147 (<0.001) | 0.9008 (<0.001) | 0.9058 (<0.001) | 0.9719 (<0.001) | 0.75 (<0.001) | 0.5746 (0.004) |
| Observations | 11015 | 11015 | 11015 | 11015 | 11015 | 11015 |
| MSA Fixed Effects | 192 | - | - | - | - | - |
| School District Fixed Effects | - | 1659 | 1659 | 1659 | 1659 | 1659 |
| Adjusted $R^2$ | 0.034 | 0.113 | 0.123 | 0.295 | 0.330 | 0.348 |

Appendix Table A6: Regression Results: MSAs 100-499 Tracts for Wilson-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.0052 (<0.001) | 0.0091 (<0.001) | 0.007 (<0.001) | 0.0012 (0.017) | 0.0293 (<0.001) | 0.0744 (<0.001) |
| Distance-Squared | - | - | - | - | -0.0008 (<0.001) | -0.004 (<0.001) |
| Distance-Cubed | - | - | - | - | - | 5.791e-05 (<0.001) |
| Access to Public Transit in 2000 | - | - | -0.3881 (<0.001) | -0.3016 (<0.001) | -0.268 (<0.001) | -0.2346 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.8149 (<0.001) | 0.7089 (<0.001) | 0.6899 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.3767 (<0.001) | 0.3269 (<0.001) | 0.3173 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.6745 (<0.001) | 0.6233 (<0.001) | 0.6093 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.2022 (<0.001) | -0.2099 (<0.001) | -0.2119 (<0.001) |
| % Buildings Built 1960-1969 | - | - | - | -0.3785 (<0.001) | -0.3355 (<0.001) | -0.3112 (<0.001) |
| % Buildings Built 1950-1959 | - | - | - | 0.0314 (0.447) | 0.0641 (0.116) | 0.0918 (0.023) |
| % Buildings Built 1940-1949 | - | - | - | -0.7933 (<0.001) | -0.6886 (<0.001) | -0.6121 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | -0.2474 (<0.001) | -0.1754 (<0.001) | -0.0702 (0.037) |
| Intercept | 1.0331 (<0.001) | 0.9671 (<0.001) | 0.9757 (<0.001) | 1.1208 (<0.001) | 1.0033 (<0.001) | 0.8544 (<0.001) |
| Observations | 25490 | 25490 | 25490 | 25490 | 25490 | 25490 |
| MSA Fixed Effects | 123 | - | - | - | - | - |
| School District Fixed Effects | - | 2785 | 2785 | 2785 | 2785 | 2785 |
| Adjusted $R^2$ | 0.046 | 0.240 | 0.272 | 0.370 | 0.385 | 0.394 |

Appendix Table A7: Regression Results: MSAs 500-999 Tracts for Wilson-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.0079 (<0.001) | 0.0114 (<0.001) | 0.008 (<0.001) | 0.0037 (<0.001) | 0.0327 (<0.001) | 0.0406 (<0.001) |
| Distance-Squared | - | - | - | - | -0.0008 (<0.001) | -0.0013 (<0.001) |
| Distance-Cubed | - | - | - | - | - | 9.153e-06 (0.044) |
| Access to Public Transit in 2000 | - | - | -0.3007 (<0.001) | -0.2798 (<0.001) | -0.2565 (<0.001) | -0.2534 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.8018 (<0.001) | 0.744 (<0.001) | 0.7414 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.5032 (<0.001) | 0.4843 (<0.001) | 0.4896 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.7008 (<0.001) | 0.6883 (<0.001) | 0.6921 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.1625 (0.001) | -0.1575 (0.002) | -0.1544 (0.002) |
| % Buildings Built 1960-1969 | - | - | - | -0.1403 (0.002) | -0.1038 (0.020) | -0.1037 (0.020) |
| % Buildings Built 1950-1959 | - | - | - | 0.1423 (0.003) | 0.1716 (<0.001) | 0.1733 (<0.001) |
| % Buildings Built 1940-1949 | - | - | - | -0.5934 (<0.001) | -0.5122 (<0.001) | -0.5065 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | 0.3171 (<0.001) | 0.4314 (<0.001) | 0.4494 (<0.001) |
| Intercept | 0.9529 (<0.001) | 0.8272 (<0.001) | 0.9341 (<0.001) | 0.9911 (<0.001) | 0.8071 (<0.001) | 0.7872 (<0.001) |
| Observations | 14029 | 14029 | 14029 | 14029 | 14029 | 14029 |
| MSA Fixed Effects | 18 | - | - | - | - | - |
| School District Fixed Effects | - | 988 | 988 | 988 | 988 | 988 |
| Adjusted $R^2$ | 0.028 | 0.306 | 0.333 | 0.396 | 0.405 | 0.405 |

Appendix Table A8: Regression Results: MSAs 1000 or More Tracts for Wilson-Distance

| | Control for Distance | Control for Local Amenities | Control for Public Transit | Control for Dwelling Age | Include Distance-Squared | Include Distance-Cubed |
|---|---|---|---|---|---|---|
| Distance | 0.0166 (<0.001) | 0.0147 (<0.001) | 0.0096 (<0.001) | 0.0082 (<0.001) | 0.017 (<0.001) | -0.0147 (<0.001) |
| Distance-Squared | - | - | - | - | -0.0003 (<0.001) | 0.0019 (<0.001) |
| Distance-Cubed | - | - | - | - | - | 4.24e-06 (<0.001) |
| Access to Public Transit in 2000 | - | - | -0.237 (<0.001) | -0.2418 (<0.001) | -0.2367 (<0.001) | -0.2398 (<0.001) |
| % Buildings Built 1999-2000 | - | - | - | 0.8707 (<0.001) | 0.877 (<0.001) | 0.84 (<0.001) |
| % Buildings Built 1995-1998 | - | - | - | 0.6484 (<0.001) | 0.6517 (<0.001) | 0.6447 (<0.001) |
| % Buildings Built 1990-1994 | - | - | - | 0.6382 (<0.001) | 0.6459 (<0.001) | 0.6464 (<0.001) |
| % Buildings Built 1970-1979 | - | - | - | -0.1077 (0.008) | -0.1077 (0.008) | -0.1017 (0.012) |
| % Buildings Built 1960-1969 | - | - | - | -0.1223 (0.001) | -0.1181 (0.001) | -0.1093 (0.002) |
| % Buildings Built 1950-1959 | - | - | - | 0.2825 (<0.001) | 0.2831 (<0.001) | 0.2941 (<0.001) |
| % Buildings Built 1940-1949 | - | - | - | -0.5787 (<0.001) | -0.5838 (<0.001) | -0.5356 (<0.001) |
| % Buildings Built 1939 or Earlier | - | - | - | 0.4035 (<0.001) | 0.437 (<0.001) | 0.4088 (<0.001) |
| Intercept | 0.804 (<0.001) | 0.8151 (0.001) | 1.0137 (<0.001) | 0.7374 (0.003) | 0.7808 (0.001) | 0.9651 (<0.001) |
| Observations | 26521 | 26521 | 26521 | 26521 | 26521 | 26521 |
| MSA Fixed Effects | 13 | - | - | - | - | - |
| School District Fixed Effects | - | 1634 | 1634 | 1634 | 1634 | 1634 |
| Adjusted $R^2$ | 0.177 | 0.437 | 0.449 | 0.483 | 0.483 | 0.485 |