# Assessing the Assessor's Assessments

Noah M. Braun, Juan M. Campanario, Steven A. Lann, Ashley E. Orr

September 18, 2022

## Abstract

This project seeks to understand if there is systematic, detectable racial or demographic bias in tax assessments of real estate properties in Allegheny County, Pennsylvania. We combine data from the Western Pennsylvania Regional Data Center with Pennsylvania Spatial Data Access and the American Community Survey of the U.S. Census. Employing a series of hedonic regressions of sale price and assessed value as a function of parcel characteristics and neighborhood characteristics, we find that inclusion of racial composition: percent non-white in a census block group only very marginally improves the predictive power of our specifications. Following Avenancio-León and Howard (2020), we compare the log ratio of assessed value to sale price and find that the housing market undervalues, and assessors overvalue, homes in areas with larger percentages of the population which is non-white, lower educated, or uses public assistance. Interestingly for the case of Pittsburgh, we find that assessors undervalue the benefit of access to public transit.

**Keywords**— Property Assessment, Pittsburgh

# 1   Project Description

In this project we seek to understand what neighborhood level socioeconomic and demographic characteristics assessors leverage in the property tax assessment of real estate in Allegheny, County, Pennsylvania. In Pennsylvania, all real estate is treated as one class for the purpose of property taxation and Article VIII, Section 1 of the Pennsylvania Constitution requires, "[a]ll taxes be uniform, upon the same class of subject...," which when put into practice implies that no taxpayer should pay more or less than their proportionate share of taxes [5]. In Allegheny County, specifically, the last county wide assessment was conducted in 2013, using 2012 as the base year, and since then individual property tax assessment are conducted by either an appeal by the taxing body, such as the school district, or by the homeowner. Tax assessments often follow the sale of a property, particularly if the sale price is significantly greater than the assessed value or after a significant improvement is made on the property which requires a building permit. Assessors when assessing the fair market value of a property employ comparisons to other similar properties within the same geographic area of the property of focus and use ratio comparisons to determine the assessed value.

To assess if Allegheny Country assessors accurately capture realized market values in their assessment, we combine data from the Western Pennsylvania Regional Data Center with Pennsylvania Spatial Data Access and the American Community Survey of the U.S. Census at the block group level. We estimate a series of hedonic models of sale price and fair market assessed value and iteratively include socioeconomic or demographic factors from the ACS to explore if/ how they might systematically influence the valuation of the property and the explanatory power of our specifications. Following Avenancio-León and Howard (2020)[1], we analyze the log ratio of assessed value to sale price (realized value) to detect whether assessors undervalue or overvalue properties in block groups conditional on their socioeconomic/ demographic composition.

Interpreting our descriptive hedonic specifications, we examine the amount of variation in sale price or assessed value which is explained using just the housing and parcel characteristics recorded in the assessor files. We include block group fixed effects to this specification as an upper bound on the additional explanatory power which the block group socioeconomic and demographic variables will contribute. We find that the bulk of the increase in $R^2$ is accounted for when we include imputed values of average parcel characteristics within a block group. Adding the percent of the block group with a high school degree or less, percent using public assistance, percent using public transit, and percent nonwhite only marginally increase our explanatory power. As anticipated, the valuation (both assessed and sale price) of properties in block groups are negatively correlated with a the percentage of the population which is on public assistance, are lower educated, and are non-white.

Examining the log ratio of the assessed value to the sale price, we find properties are overvalued in block groups with a higher percentage of the population which is lower educated, on public assistance, and non-white. This is consistent with the literature as the market (via sale price) punishes these neighborhood characteristics. Interestingly, in the context of Pittsburgh, assessors undervalue relative to the market, the importance of public transit and means of other transport to work, such as walking on foot- a finding which is consistent anecdotally with Pittsburgh's high use of public transit and cultural importance of neighborhoods which are walkable and accessible for bike commutes.

# 2   Data

To identify discrepancies between the fair market value and the sale prices of houses with regards to the race, we utilize both the 2013 Allegheny County Property Assessments[4] and property sales data[3] from the Western Pennsylvania Regional Data Center. The property assessments contain information on the value and size of the land, previous sales, tax, municipality, and neighborhood codes, and building characteristics at the parcel level. The building characteristics include the housing grades, the year built, living area in square footage, the number of bedrooms, bathrooms, and fireplaces, data on air-conditioning, as well as other variables pertaining to the quality of construction and available space. The data differentiates between residential, utilities, industrial, commercial, government, and

agricultural buildings as well as sales descriptions, i.e. love and affection, sheriff sales, estate sales, valid sales, etc. For analyzing the assessor's fair market valuations, the sample is restricted to residential properties. The sample is then further restricted to valid sales, or arm's length sales, i.e. sales the Allegheny County Office of Property Assessments deemed to be representative of the current market value. Lastly, we focus on houses sold after 2010 to avoid extrapolating prices over an extend time period and to remove any unusual sales or purchases conducted as a response to the 2008-2009 financial crisis.

Ideally, analysis would use demographic characteristics at the neighborhood level, which is defined as a "geographic area exhibiting a high degree of homogeneity in economic amenities, land use, economic trends, and property characteristics such as quality, age, and condition," (OPA). By doing so, one could compare valuations and sales for very similar neighborhoods that vary along race. As detailed demographic data is not available at that level of granularity, we use block-group level from the American Community Survey[2]. We use percentages for high school degree attainment, population age ranges, public assistance program use, public transit use, and the use of other forms of transportation. We construct the percentage of non-White population by summing over the percentages for Black, Indigenous, Hispanic, and other people. To combine the demographic data to the property assessments data, the assessments are first merged to the 2020 Allegheny County Parcel map provided by the Pennsylvania Geospatial Data Clearinghouse, which contains the block-group values for each parcel. We also construct average parcel characteristics at the block-group level utilizing the full sample of assessments to create more accurate average parcel characteristic data.

In Table 1, we constructed average measures of parcel and demographic characteristics for the sales data for each quartile of percent non-White. Parcels sold in between the $25^{th}$ and $75^{th}$ percentile of non-White are on average sold at a higher price, have larger living spaces, and are located in areas with relatively younger populations.[1] In Table 2, we compare the explained variation in fair market valuations and sale prices when controlling for parcel characteristics alone and with neighborhood code and block-group fixed effects. While all specifications capture a larger portion of the variance in both outcomes, it is clear the neighborhood fixed effects would be preferred.

# 3    Methods

To address these hypotheses, we regress both the assessment values and realized sales values on property and neighborhood characteristics. We estimate the following equations,

$$\text{Assessed Value}_i = \beta_0 + \beta_1 \mathbf{X_i} + \beta_2 \mathbf{N_j} + \epsilon_i \tag{1}$$

$$\text{Sales Value}_i = \alpha_0 + \alpha_1 \mathbf{X_i} + \alpha_2 \mathbf{N_j} + \delta_t + \epsilon_i \tag{2}$$

---

[1]See Appendix for fair market valuation summary statistics.

Table 1: Summary Statistics: Housing and Demographic Characteristics

|  | (1)<br>$0 - 25^{\text{th}}$ | (2)<br>$25 - 50^{\text{th}}$ | (3)<br>$50 - 75^{\text{th}}$ | (4)<br>$75 - 100^{\text{th}}$ |
|---|---|---|---|---|
| Fair Market | 160236.144 | 193834.743 | 184124.240 | 118580.862 |
|  | (106963.111) | (157626.476) | (165180.480) | (118197.474) |
| Sale | 178768.775 | 215102.275 | 204419.910 | 132289.974 |
|  | (118205.458) | (169546.972) | (179425.419) | (132659.079) |
| Ratio of Fair Mkt and Sale | 0.931 | 0.931 | 0.930 | 0.945 |
|  | (0.407) | (0.404) | (0.388) | (0.380) |
| Living Area (ft$^2$) | 1640.693 | 1847.578 | 1798.788 | 1615.873 |
|  | (704.697) | (891.409) | (903.868) | (720.902) |
| Land Area (ft$^2$) | 14548.811 | 16493.049 | 13387.992 | 7322.846 |
|  | (33018.960) | (51952.492) | (33221.304) | (12923.900) |
| # of Stories | 1.540 | 1.637 | 1.659 | 1.691 |
|  | (0.491) | (0.484) | (0.497) | (0.508) |
| Physical Condition | 3.914 | 3.867 | 3.861 | 4.061 |
|  | (0.521) | (0.583) | (0.623) | (0.682) |
| # of Rooms | 6.334 | 6.671 | 6.575 | 6.333 |
|  | (1.522) | (1.813) | (1.841) | (1.791) |
| # of Bedrooms | 2.974 | 3.117 | 3.069 | 2.954 |
|  | (0.784) | (0.865) | (0.922) | (0.978) |
| # of Full Bathrooms | 1.514 | 1.659 | 1.618 | 1.425 |
|  | (0.643) | (0.744) | (0.750) | (0.649) |
| # of Fireplaces | 0.401 | 0.424 | 0.377 | 0.276 |
|  | (0.543) | (0.569) | (0.571) | (0.511) |
| Air-Conditioning | 0.764 | 0.752 | 0.716 | 0.548 |
|  | (0.425) | (0.432) | (0.451) | (0.498) |
| Year Built | 1960.241 | 1961.111 | 1956.072 | 1943.341 |
|  | (27.557) | (31.666) | (33.091) | (30.411) |
| Sale Year | 2015.048 | 2015.063 | 2015.044 | 2015.189 |
|  | (2.967) | (2.986) | (3.003) | (3.008) |
| % Non-White | 0.024 | 0.077 | 0.167 | 0.431 |
|  | (0.017) | (0.017) | (0.037) | (0.182) |
| % High School Degree | 0.329 | 0.278 | 0.281 | 0.353 |
|  | (0.130) | (0.143) | (0.139) | (0.161) |
| % Public Assistance | 0.018 | 0.015 | 0.020 | 0.043 |
|  | (0.025) | (0.024) | (0.030) | (0.060) |
| % Public Transit | 0.050 | 0.055 | 0.072 | 0.142 |
|  | (0.048) | (0.061) | (0.071) | (0.122) |
| % Under 18 | 0.183 | 0.201 | 0.191 | 0.188 |
|  | (0.061) | (0.069) | (0.078) | (0.087) |
| % 18 to 20 | 0.053 | 0.061 | 0.075 | 0.099 |
|  | (0.033) | (0.042) | (0.077) | (0.099) |
| % 45 to 60 | 0.308 | 0.300 | 0.278 | 0.257 |
|  | (0.058) | (0.063) | (0.071) | (0.085) |
| % 65+ | 0.217 | 0.184 | 0.180 | 0.178 |
|  | (0.082) | (0.072) | (0.078) | (0.083) |
| Observations | 13045 | 13836 | 12952 | 9297 |

Table 2: Comparison of Block-Group and Neighborhood Fixed Effects

|  | (1) Fair Mkt | (2) Fair Mkt | (3) Fair Mkt | (4) Sale | (5) Sale | (6) Sale |
|---|---|---|---|---|---|---|
| Housing Characteristics | Yes | Yes | Yes | Yes | Yes | Yes |
| Block Group FE | No | Yes | No | No | Yes | No |
| Neighborhood FE | No | No | Yes | No | No | Yes |
| Adjusted R-Squared | 0.708 | 0.849 | 0.898 | 0.659 | 0.824 | 0.855 |
| Observations | 440855 | 440855 | 440855 | 48975 | 48975 | 48975 |
| Clusters | 175 | 175 | 175 | 170 | 170 | 170 |

where $\mathbf{X_i}$ is parcel characteristics, including: acreage, property valuation, grade, condition, number of rooms, number of stories, stories, and square footage. We then iteratively add neighborhood characteristics, $N_j$, which include: demographic variables, use of public assistance, means of transport to work, and housing quality. We compare the unexplained variation before and after adding specific variables to understand the importance of race, age, education, and income composition of the neighbourhood. For sales value, we also include a time fixed effect, $\delta_t$, for the year of sale.

For a direct comparison of how neighborhood characteristics are valued by tax assessors and the market, we follow Avenancio-León and Howard (2020) by placing the logged ratio of assessed value to sales value on the left hand side. Specifically, we estimate the following equation,

$$\ln\left(\frac{\text{Assessed Value}_i}{\text{Sales Value}_i}\right) = \beta_0 + \beta_1 \mathbf{X_i} + \beta_2 \mathbf{N_j} + \delta_t + \epsilon_i \tag{3}$$

where $\mathbf{X_i}$ is again parcel characteristics, $N_j$ is neighborhood characteristics, and $\delta_t$ is a time fixed effect for the year of the sale. In this regression, a coefficient with a positive (negative) sign will indicate that a block-group with a larger value for the given variable is over (under) assessed relative to the market. Lastly, we run equation 3 with a neighborhood fixed effect in place of our parcel and neighborhood characteristics.

# 4 Results

The results from Specification 1 are presented in Table 3. In all tables presented, Column 1 reports the baseline specification where only parcel characteristics from the assessor files are included as independent variables and shows the variance in the assessed value explained by only the individual characteristics of the property. Column 2 reports the specification with the addition of block group fixed effects, we consider this to be an upper bound on the additional explanatory power which the block group socioeconomic and demographic

variables will contribute. In Column 3 we instead include imputed neighborhood quality by averaging the characteristics of parcels within the same block group as recorded in the assessors files. Starting with Column 4, we iteratively add in socioeconomic and demographic composition of the block groups; Column 5 additionally includes the percentage of the block group population which is non-white.

Table 3: Log Fair Market Valuation with Block-Group Demograhic Characteristics

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| % High School Degree |  |  |  | -0.794*** | -0.754*** |
|  |  |  |  | (0.103) | (0.092) |
| % Public Assistance |  |  |  | -0.647*** | -0.445*** |
|  |  |  |  | (0.145) | (0.134) |
| % Public Transit |  |  |  | -0.225** | 0.081 |
|  |  |  |  | (0.100) | (0.098) |
| % Other Transportation |  |  |  | 0.280** | 0.360*** |
|  |  |  |  | (0.133) | (0.129) |
| % Non-White |  |  |  |  | -0.482*** |
|  |  |  |  |  | (0.048) |
| Housing Characteristics | Yes | Yes | Yes | Yes | Yes |
| Block Group FE | No | Yes | No | No | No |
| Average Parcel Characteristics | No | No | Yes | Yes | Yes |
| ACS Demographics | No | No | No | Yes | Yes |
| Adjusted R-Squared | 0.708 | 0.849 | 0.805 | 0.821 | 0.827 |
| Observations | 440855 | 440855 | 440855 | 440822 | 440822 |
| Clusters | 175 | 175 | 175 | 175 | 175 |

$^{*}$ $p < .1$, $^{**}$ $p < .05$, $^{***}$ $p < .01$

We find that the bulk of the increase in $R^2$ is accounted for when we include imputed values of average parcel characteristics within a block group. Adding the percent of the block group with a high school degree or less, percent using public assistance, percent using public transit, and percent nonwhite only marginally increase our explanatory power. As anticipated, the valuation (both assessed and sale price) of properties in block groups are negatively correlated with a the percentage of the population which is on public assistance, are lower educated, and are non-white.

The results from Specification 2, are presented in Table 4. Many of the results from Table 3 still hold here; the coefficients on percentage non-white, percentage on public assistance, and percentage with a high school degree or less are all significantly negative. In addition,

the share of households that use public transportation is positively associated with sales price, implying that the market values greater access to public transit and other alternative modes of transportation.

Table 4: Log Sale Price with Block-Group Demograhic Characteristics

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| % High School Degree |  |  |  | -0.889*** | -0.852*** |
|  |  |  |  | (0.112) | (0.101) |
| % Public Assistance |  |  |  | -0.928*** | -0.658*** |
|  |  |  |  | (0.175) | (0.147) |
| % Public Transit |  |  |  | -0.088 | 0.246** |
|  |  |  |  | (0.114) | (0.101) |
| % Other Transportation |  |  |  | 0.534*** | 0.634*** |
|  |  |  |  | (0.147) | (0.147) |
| % Non-White |  |  |  |  | -0.582*** |
|  |  |  |  |  | (0.050) |
| Housing Characteristics | Yes | Yes | Yes | Yes | Yes |
| Block Group FE | No | Yes | No | No | No |
| Average Parcel Characteristics | No | No | Yes | Yes | Yes |
| ACS Demographics | No | No | No | Yes | Yes |
| Adjusted R-Squared | 0.659 | 0.824 | 0.762 | 0.785 | 0.794 |
| Observations | 48975 | 48975 | 48975 | 48972 | 48972 |
| Clusters | 170 | 170 | 170 | 169 | 169 |

* $p < .1$, ** $p < .05$, *** $p < .01$

Examining the log ratio of the assessed value to the sale price, as in Specification 3, we report results in Table 5. Interpreting this specification, recall that a positive coefficient indicates that as that characteristic increases, the fair market value increases more than the sales price.

Here we find properties are overvalued by the assessors for block groups with a higher percentage of the population which is lower educated, on public assistance, and non-white. This is consistent with the literature as the market (via sale price) is recorded as punishing these neighborhood characteristics. One plausible mechanism which could rationalize this result is that assessors are intentionally not considering demographic neighborhood composition even though these neighborhood characteristics are influencing sale price in the housing market.

Interestingly, in the context of Pittsburgh, assessors undervalue relative to the market the importance of public transit and means of other transport to work, such as biking or

walking on foot. Recall, in Specification 2, we similarly find that public transit and other means of transportation are positively and significantly correlated with the parcel's sale price. This finding is consistent anecdotally with Pittsburgh's high use of public transit and cultural importance of neighborhoods which are walkable and accessible for bike commuters.

Table 5: Log Ratio of Fair Market Valuation to Sale Price

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| % High School Degree | | | | 0.076*** | 0.070*** |
| | | | | (0.018) | (0.017) |
| % Public Assistance | | | | 0.116*** | 0.079* |
| | | | | (0.041) | (0.041) |
| % Public Transit | | | | -0.057** | -0.103*** |
| | | | | (0.027) | (0.024) |
| % Other Transportation | | | | -0.078** | -0.091*** |
| | | | | (0.034) | (0.034) |
| % Non-White | | | | | 0.080*** |
| | | | | | (0.012) |
| Housing Characteristics | Yes | Yes | Yes | Yes | Yes |
| Block Group FE | No | Yes | No | No | No |
| Average Parcel Characteristics | No | No | Yes | Yes | Yes |
| ACS Demographics | No | No | No | Yes | Yes |
| Adjusted R-Squared | 0.114 | 0.149 | 0.124 | 0.127 | 0.129 |
| Observations | 48975 | 48975 | 48975 | 48972 | 48972 |
| Clusters | 170 | 170 | 170 | 169 | 169 |

* $p < .1$, ** $p < .05$, *** $p < .01$

As a final exercise, we plot on a map of Allegheny county the coefficients on the neighborhood fixed effects from our Specification 3. The neighborhoods shaded darker are areas with high assessments relative to sales prices. We draw attention to two areas on the map. Area (1), containing the neighborhoods of Polish Hill, Middle Hill, and Bedford Dwellings, have high assessed values relative to sales prices. These neighborhoods have some of the highest shares of non-white residents in Allegheny county, with over 80% of residents being Black. This is contrasted by area (2), the neighborhood of Mount Lebanon. One of the most affluent suburbs of Pittsburgh, it also has some of the lowest relative assessed values in the county. The comparison of these highlighted neighborhoods shows qualitatively the disparities in relative assessed value by racial composition and socioeconomic status.
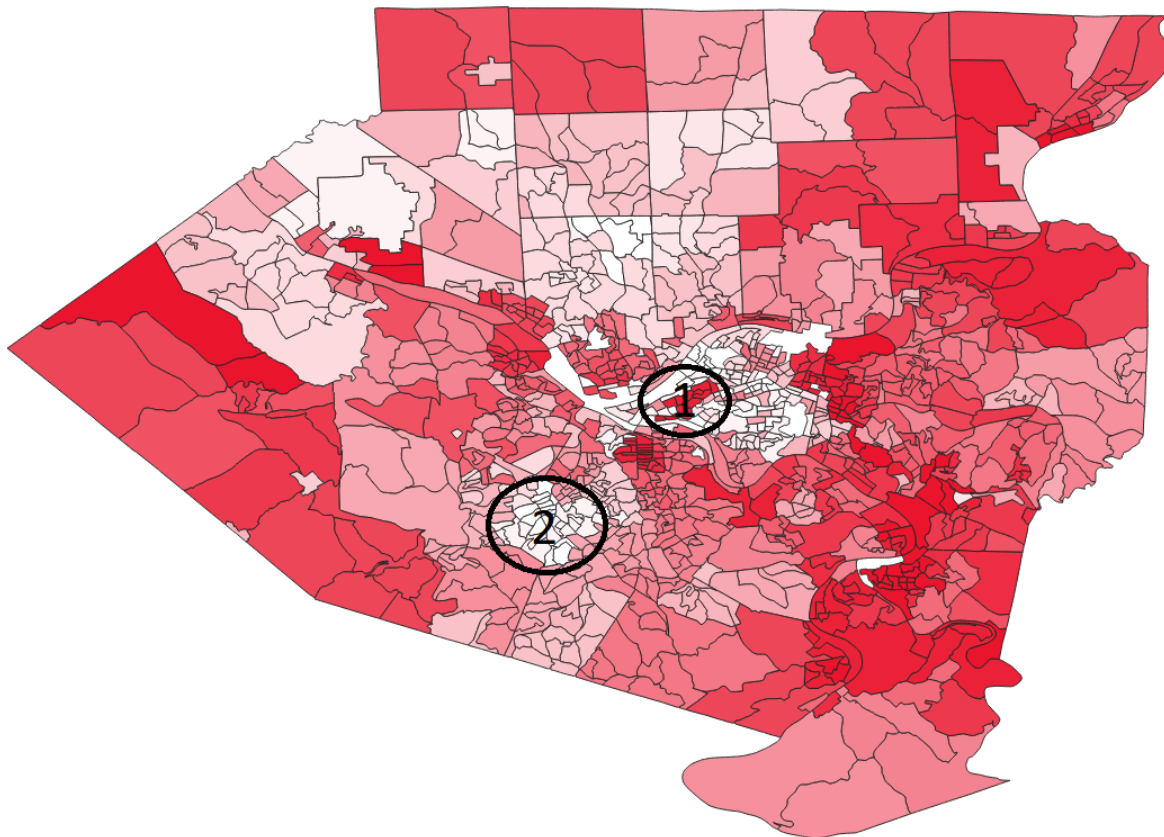
Figure 1: Darker areas represent higher assessed values relative to sale prices.

# References

[1] Carlos Avenancio-Leon and Troup Howard. The assessment gap: Racial inequalities in property taxation. June 2020.

[2] United States Census Bureau. American Community Survey Data [dataset], 2020. https://www.census.gov/programs-surveys/acs/data.html.

[3] Western Pennsylvania Regional Data Center. Allegheny County Property Sale Transactions [dataset], 2013- October 2020. https://data.wprdc.org/dataset/real-estate-sales.

[4] Western Pennsylvania Regional Data Center. Parcel Centroids in Allegheny County with Geographic Identifiers [dataset], December 2019. https://data.wprdc.org/dataset/parcel-centroids-in-allegheny-county-with-geographic-identifiers.

[5] Local Government Commission Property Assessment Reform Task Force. Pennsylvania Property Assessment: A Self-Evaluation Guide for County Officials. June 13, 2018.

# 5 Appendix

Table 6: Summary Statistics: Housing and Demographic Characteristics

|  | (1) $0 - 25^{th}$ | (2) $25 - 50^{th}$ | (3) $50 - 75^{th}$ | (4) $75 - 100^{th}$ |
|---|---|---|---|---|
| Fair Market | 147070.847 | 168830.543 | 158895.639 | 89684.096 |
|  | (99870.803) | (147645.121) | (154630.616) | (101529.162) |
| Sale | 110616.283 | 132030.055 | 126679.516 | 83005.972 |
|  | (154990.813) | (281974.226) | (178069.280) | (324613.528) |
| Ratio of Fair Mkt and Sale | 26685.808 | 27962.679 | 26099.393 | 15613.606 |
|  | (63052.684) | (80638.429) | (78481.246) | (49194.128) |
| Living Area (ft$^2$) | 1642.941 | 1794.900 | 1771.051 | 1621.079 |
|  | (693.591) | (863.772) | (883.118) | (725.210) |
| Land Area (ft$^2$) | 21290.708 | 18655.868 | 14789.049 | 6793.707 |
|  | (102518.100) | (67109.603) | (67242.734) | (19977.394) |
| # of Stories | 1.525 | 1.627 | 1.673 | 1.778 |
|  | (0.494) | (0.491) | (0.500) | (0.484) |
| Physical Condition | 3.994 | 3.972 | 3.999 | 4.324 |
|  | (0.564) | (0.620) | (0.686) | (0.810) |
| # of Rooms | 6.354 | 6.617 | 6.575 | 6.403 |
|  | (1.501) | (1.746) | (1.833) | (1.905) |
| # of Bedrooms | 2.983 | 3.090 | 3.068 | 2.996 |
|  | (0.787) | (0.871) | (0.946) | (1.014) |
| # of Full Bathrooms | 1.469 | 1.566 | 1.546 | 1.362 |
|  | (0.622) | (0.710) | (0.735) | (0.634) |
| # of Fireplaces | 0.393 | 0.415 | 0.360 | 0.226 |
|  | (0.538) | (0.569) | (0.558) | (0.475) |
| Air-Conditioning | 0.693 | 0.665 | 0.607 | 0.379 |
|  | (0.461) | (0.472) | (0.488) | (0.485) |
| Year Built | 1955.849 | 1954.554 | 1948.317 | 1933.842 |
|  | (29.332) | (32.453) | (33.613) | (29.509) |
| Sale Year | 2005.636 | 2006.125 | 2006.443 | 2006.031 |
|  | (11.851) | (11.622) | (11.949) | (11.570) |
| % Non-White | 0.024 | 0.078 | 0.169 | 0.502 |
|  | (0.017) | (0.017) | (0.037) | (0.222) |
| % High School Degree | 0.344 | 0.299 | 0.303 | 0.397 |
|  | (0.130) | (0.150) | (0.145) | (0.165) |
| % Public Assistance | 0.020 | 0.018 | 0.022 | 0.058 |
|  | (0.027) | (0.026) | (0.032) | (0.070) |
| % Public Transit | 0.051 | 0.058 | 0.079 | 0.171 |
|  | (0.051) | (0.062) | (0.077) | (0.141) |
| % Under 18 | 0.180 | 0.194 | 0.187 | 0.199 |
|  | (0.061) | (0.069) | (0.076) | (0.093) |
| % 18 to 20 | 0.055 | 0.064 | 0.079 | 0.099 |
|  | (0.032) | (0.053) | (0.084) | (0.094) |
| % 45 to 60 | 0.310 | 0.299 | 0.278 | 0.257 |
|  | (0.059) | (0.065) | (0.074) | (0.085) |
| % 65+ | 0.216 | 0.185 | 0.178 | 0.173 |
|  | (0.080) | (0.072) | (0.075) | (0.083) |
| Observations | 98287 | 100746 | 100829 | 99979 |