Evaluating Urban Connectivity and its Relationship with Property Appraisal Values in Allegheny County

May 6th, 2024

MSc Real Estate Studies - Faculty of Spatial Sciences - University of Groningen

Supervisor Mark van Duijn

Master Thesis of Jolevi van Delft - s3896234

Abstract

This thesis aims to quantify the relationship between connectivity features, such as the density of roads (links) and intersections (nodes), and property appraisal values in Allegheny County.

By using a hedonic pricing model, the research analyzes data from the Allegheny County Property Transactions Database and the National Neighborhood Data Archive on Urban Connectivity. The findings show that both link and node densities are positively related to higher property values, showing the importance of well-planned transport infrastructure in enhancing property desirability and economic value.

Furthermore, the study explores the heterogeneity in connectivity relationships across different neighborhood densities. Results show that urban connectivity has a more substantial relationship with property appraisal values in less densely connected outer-city areas compared to more densely connected inner-city regions. This suggests that the benefits of improved connectivity are more substantial in suburban or rural areas with lower density, while excessive infrastructure in dense urban areas can lead to negative effects that reduce property fair market values.

This thesis offers insights for policymakers and urban planners by showing the need for strategic infrastructure development that takes local neighborhood characteristics into account, such as neighborhood density, existing transport networks, and housing attributes, aiming to improve urban growth and maximizing the benefits of connectivity improvements.

1. Introduction

1.1 Motivation

As we witness a global surge in urbanization, the role of urban connectivity has become increasingly crucial. It is not just physical networks that are important, but also the growing digital connections that are reshaping the way we live and work. This trend is particularly impactful in the real estate market, where enhanced connectivity or accessibility, either through improved transport systems or advanced digital infrastructure, is influencing property values. Improved connectivity anecdotally attracts businesses and residents, leading to increased demand for properties in well-connected areas. Next to this, enhanced transport and digital infrastructure contribute to a better quality of life, making those areas more desirable and increasing real estate prices.

Media reports underline these findings. Pressley (2023) discussed the ways infrastructure enhancements can increase property market values, and Shaw (2022) observed that urban areas with strong connectivity typically experience more rapid property price growth compared to less connected or remote areas. Recently, Mischke et al. (2023) mentioned that the pandemic has changed real estate in lasting ways, emphasizing the need to understand how connectivity and property values relate in these transformative times. With these observations in mind, it is clear there is a continued need to understand the role of connectivity in the real estate landscape.

Allegheny County is a unique case for such an investigation. Historically, Allegheny County is known for its steel production, and as the county has developed over the years it has evolved into a labor market that is mostly dominated by healthcare, education, and technology sectors. Major hubs like the University of Pittsburgh and UPMC, along with tech companies, define where people work and how they commute (Wikimedia Foundation, 2023a). Infrastructure development and maintenance in Allegheny County are also of great significance. Because there are many rivers and bridges, transportation and connectivity have always been important to the region's growth and functionality. Over the past two decades, Allegheny County has seen some significant infrastructure projects. Projects such as the North Shore Connector expanded the rail network, allowing for easier access between North Shore and downtown Pittsburgh (Wikimedia Foundation, 2023b). These projects do not only improve commuting but can also have a considerable impact on property values in nearby areas.

Budget for infrastructure in Allegheny County has always been a point of discussion among policymakers. By recognizing the important role that connectivity plays in the economic and social aspects of the region, much has been done to maintain and enhance the county's infrastructure. This emphasis on infrastructure in Allegheny County offers a solid perspective for examining the correlation between urban connectivity and property values. By focusing on this county, this research can provide insights into the societal and economic realities of a region that has undergone significant transformation and continues to evolve. Understanding these dynamics can offer valuable lessons for policymakers and other regions undergoing similar transitions and developments.

1.2 Academic Relevance

Building on the idea of urban connectivity, much research investigates how it influences property values. Historically, this general relationship has been acknowledged, but as urban environments evolve, it is important to revisit and update our understanding.

Studies by Ghosh et al. (2021), Deng et al. (2019), Bujanda & Fullerton (2017), and Cohen & Schaffner (2021) consistently find that investments in infrastructure typically increase nearby property values, demonstrating the significant influence of physical connectivity.

The theoretical explanation for this relationship depends on increased accessibility and reduced travel costs, which improves the utility and desirability of properties. As shown in Chapter 2, this improved accessibility leads to higher demand for properties in well-connected areas against a fixed supply of land, thereby increasing property values (Evans, 2008). This concept is supported by the bid rent theory, which argues that properties closer to urban centers or significant infrastructure receive higher prices due to their time and travel cost savings (Alonso, 1964).

As existing literature suggests there is more to learn from localized studies (Ghosh et al., 2021; Deng et al., 2019) due to the vast variability found within each city, this thesis seeks to contribute by focusing on Allegheny County, a region with a unique blend of historical and contemporary infrastructure. While some aspects of past studies are relevant, every city is different, meaning that not every conclusion is universally applicable. To build on these insights, this study aims to examine the relationship between urban connectivity and property fair market values, as well as explore the variations across different neighborhood densities within the county.

Unlike broader studies that often focus on one specific infrastructure project in one location or the effects of infrastructure improvements over time, this research investigates the entire county by measuring infrastructure in a different way. This allows for heterogeneity analyses to understand different impacts across various neighborhoods. The research not only investigates how specific physical connectivity features, such as roads (links) and intersections (nodes), relate to property appraisal values but also compares this relationship between more densely connected inner-city areas and those in less densely connected outer-city areas. This addresses a notable gap in existing research, which often provides a broader overview without delving into distinct infrastructure presence or its relationship across various urban settings.

Understanding both the general relationship of connectivity and the variations across neighborhoods can offer valuable insights for urban planners and policymakers. By identifying how infrastructure improvements associate with property values, this study can help guide interventions that maximize benefits and minimize unintended consequences and inform effective urban development strategies.

1.3 Research Problem Statement

The research aim of this study is to investigate how urban connectivity relates to fair market values in Allegheny County.

Central Research Question:

How does urban connectivity associate with property appraisal values in Allegheny County?

Sub-questions:

1. How do contemporary theories on urban development and connectivity conceptualize the relationship between transportation and property values?

This question will be explored by reviewing existing research and theories from academic journals. Looking into how urban development and connectivity are thought to affect property values, with a focus on the measurable relationship between transportation improvements and property values.

2. How do specific attributes of urban connectivity, such as infrastructure features, associate with property appraisal values in Allegheny County?

To investigate how specific attributes of urban connectivity, such as infrastructure features, contribute to property appraisal values in Allegheny County, the study will use multivariate regression analyses on property data and urban connectivity metrics. Additionally, GIS maps will be used to visually analyze and present the spatial distribution of these connectivity attributes and their correlation with property appraisal values. These methods will provide a clear picture of how transportation infrastructure relates to real estate fair market values across the county.

3. To what extent does the relationship between urban connectivity and property appraisal values differ between densely connected inner-city areas and less densely connected outer-city areas in Allegheny County?

This sub-question aims to explore how urban connectivity relates to property appraisal values across neighborhoods with different urban densities. By employing a Chow F Test, the analysis will compare neighborhoods classified by their link/node ratios into more densely connected (inner-city) and less densely connected (outer-city) areas.

1.4 Outline

Following the research problem statement, the thesis will progress through chapters covering the literature review, outlining existing theories and gaps, data and methods, describing the analytical approach using a visual analysis, results and discussion, presenting findings on urban connectivity's relationship with property appraisal values and how they relate to the literature, and concluding with a conclusion that summarizes and interprets these findings while offering future research ideas.

2. Literature Review

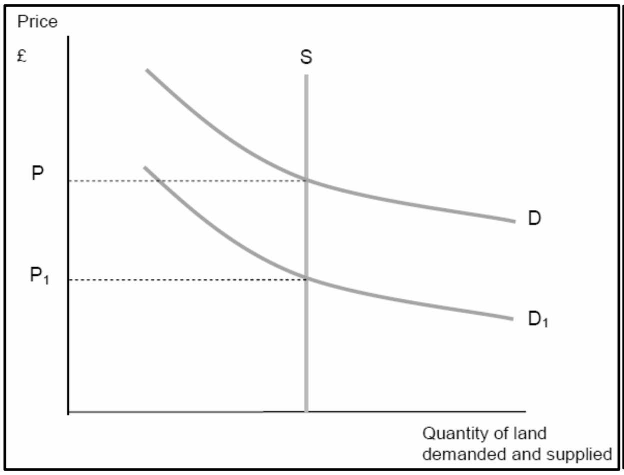
The dynamic between a city's transport infrastructure and the value of properties in the area is a widely discussed area of urban economics, particularly as we see cities growing and becoming more connected. Currently, it is generally accepted that a place's accessibility, due to roads, bridges, and other connectivity features can drive up the demand for homes there, leading to an increase in how much these homes are worth (Bujanda & Fullerton, 2017). Research consistently shows a positive relationship between infrastructure accessibility and housing prices. Accessibility, defined as the ease with which residents can reach key locations such as city centers, workplaces, educational institutions, and public transportation stops, plays a significant role in determining housing prices (Demirdağ, 2023)

Empirical evidence supports the positive impact of transportation infrastructure on real estate values. For instance, the introduction of the Madrid Metro Line 12 (Metrosur) in Spain showed that better accessibility to metro stations positively influenced housing prices. Similarly, in Seoul, South Korea, better local and systemwide accessibility was associated with higher apartment sales prices (Shin, 2007). The type of transportation infrastructure also matters. Research has shown that rail, bus rapid transit, and even conventional bus transit systems can affect property prices. In Xiamen, China, both accessibility to bus stops and travel time to city centers significantly influenced housing prices, with bus frequency having a larger impact in peripheral areas compared to central ones (Yang, 2020).

The relationship between housing prices and accessibility is varied, involving various types of transportation infrastructure and showing spatial variations. Both current and past investments in transportation infrastructure can have significant local market effects, affecting housing prices on an individual basis and across broader areas (Mikelbank, 2000). The consistent finding across different contexts and methodologies is that improved accessibility generally leads to higher property values. In the context of major infrastructure developments, such as the construction of the Westerschelde Tunnel in the Netherlands, the impact of accessibility on housing prices is similar. Hoogendoorn (2019) provides evidence from this infrastructure project, indicating a strong positive relationship between accessibility and house prices, with effects even before the actual opening of the tunnel.

However, not all studies show a uniform positive impact. Adair (2000) examined the Belfast Urban Area and found that while accessibility had little significance in explaining house price variations at a city-wide scale, it was a more important factor in lower-income sub-markets. This shows the importance of considering sub-markets when analyzing the relationship between accessibility and housing prices. However, these study results do not fully explain the underlying drivers that are at hand here.

To more extensively explore the relationship between city transportation infrastructure and property values, it is important to begin with fundamental supply and demand concepts, drawing on insights from Evans (2008) and Alonso (1964). The demand for properties in accessible areas is primarily driven by the convenience they offer, allowing residents to save on travel costs and time. This concept is shown by the bid rent theory, which states that properties closer to city centers or major transport links get higher prices due to their reduced commuting costs. The increased property values near city centers and transport hubs show the higher price that people are willing to pay for this convenience and accessibility. On the supply side, the availability of land for development plays an important role. Even as demand increases due to improved accessibility, the fixed supply of land means that only so much development can occur, which naturally leads to higher property prices in these desirable areas. The principle here is simple: more people want to live where they can easily get around, and there is only so much land that can be developed, leading to higher prices in these areas.



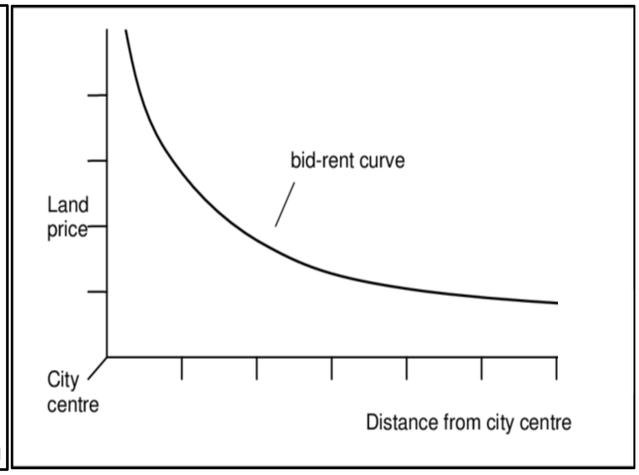


Figure 1: Land Supply & Demand Graph by Evans   Figure 2: Bid Rent Curve Theory by William Alonso

As we look at how these factors play out, the two figures above help visualize these concepts. Figure 1 shows the effect of increased demand for properties in accessible areas on property prices, showing the interaction between supply and demand. When there are positive changes in accessibility, such as improved transportation infrastructure or better connectivity, the demand curve (D_1) in Figure 1 shifts upwards (D), indicating an increase in demand for properties in these well-connected areas. This leads to higher property prices (P), as more buyers are willing to pay more for the benefits of enhanced accessibility.

Figure 2 illustrates the bid rent theory, showing how property values change depending on their location relative to the city center or transport hubs. Positive changes in accessibility, which lead to reduced transport costs, cause the bid rent curve to flatten out. This flattening happens because lower transportation costs reduce the extra amount buyers are willing to pay for properties closer to the city center or transport hubs. As a result, property values in areas further from these central points increase, showing the reduced cost and time of commuting. These visuals provide a graphical representation of how accessibility influences real estate values across different areas.

While saving on travel costs and time, as highlighted by the bid-rent theory and supply and demand concepts, leads to an increase in demand, the importance of other factors that influence demand must also be recognized. These mechanics provide valuable insights into property value variations, but infrastructure is not the only influential factor. The decision-making process of potential homeowners also plays a significant role. Studies in the field of housing dynamics, such as the works of Thang (2001) & Margulis (1988), indicate that choices regarding housing (re)location are often shaped by various personal and economic factors, including age, family size, education, employment type, and income. Additionally, the convenience of access to amenities also plays a big role in these decisions (Peris & Enault, 2023). This aspect is particularly important for individuals when determining their future housing arrangements. Areas that offer a desirable mix of accessible amenities tend to attract the most people, which is what then results in corresponding differences in house prices in these regions. This aligns with Haugen (2011), who highlighted the significance of proximity to work, services, and social activities in homebuyer decision-making. This increased demand influences the housing prices in these regions. The literature suggests that the relationship between infrastructure and housing prices is not a direct, linear relationship. Instead, it is shaped by a combination of individual preferences, economic conditions, and the availability of amenities, all of which are important in determining the desirability and value of these properties.

More practical studies back up these findings, showing that when infrastructure and therefore access to certain amenities improves, so do the prices of nearby homes. Take, for example, the mentioned increase in home values seen when a new highway is built (Cohen & Schaffner's, 2021). In further examining the relationship between specific connectivity features and property values, the study by Yiu & Wong (2005), is also relevant. This research explores how even anticipated improvements in physical connectivity, such as new transportation projects, influence housing prices. The findings suggest that the expectation of enhanced connectivity through infrastructure development, particularly the construction of tunnels, can elevate property values before the completion of these projects.

But even though there is a generally acknowledged link between connectivity features and higher home prices, it is not always clear how much specific aspects of a city's infrastructure relate to these prices. The hedonic pricing theory, which states that the price of a home is influenced by many different features, including location, can help explain the relevance of such features. This model is key to understanding how particular aspects of infrastructure relate to home values. These observations lay the basis for the first hypothesis:

> **H1:** Specific attributes of urban connectivity, such as the number of links and nodes of transportation infrastructure, have a significant and positive relationship to property appraisal values.

The effect of neighborhood differences is another aspect to consider. It is possible that the relationship between improved urban connections and home prices is not equal across different regions within Allegheny County. Differences in local economic activities, demographic composition, and neighborhood characteristics could also lead to different associations of the same infrastructural enhancements in different areas.

As discussed before, the literature on urban economics suggests that the relationship between infrastructure improvements and property values is complex and influenced by local contexts (Adair, 2000; Mikelbank, 2000). For instance, Mathur (2008) provides evidence from a study conducted in King County, Washington, showing that the effects of infrastructure and services on property values vary significantly depending on the quality and age of housing, as well as the existing urban area of the neighborhoods. This shows that the benefits of improved connectivity are not felt the same everywhere but are influenced by specific neighborhood properties.

Furthermore, the bid rent theory (Alonso, 1964) suggests that improvements in connectivity may have a greater impact on property values in outer city areas than in inner city areas. Positive changes in accessibility reduce transportation costs, which in turn decreases the premium buyers are willing to pay for properties closer to the city center or transport hubs. As a result, property values in areas further from these central points increase, showing the reduced cost and time of commuting. This theory implies that outer-city areas might benefit more from improvements in connectivity, as these changes make them more attractive relative to the previously more desirable inner-city areas.

This leads us to the second hypothesis, exploring the potentially diverse relationship of urban connectivity across varying densities within the county.

> **H2:** The relationship of urban connectivity to property appraisal values is significant and greater in less densely connected inner-city areas compared to more densely connected outer-city areas.

3. Methods & Data

3.1 Methodology

The methodology utilizes a hedonic pricing model using a multiple linear regression in Stata. This model is based on the hedonic pricing theory, which states that the price of a good is determined by the characteristics and the benefits it provides to consumers. Within the context of this study, hedonic pricing theory suggests that property appraisal values are related to attributes, such as location, size, number of rooms, age, and proximity to infrastructure and amenities. By analyzing these factors, the model aims to isolate the specific relationship between urban connectivity and property fair market values. This results in the following regression equation:

$$\begin{aligned} log(FairMarketTotal) &= \beta 0 + \beta 1 \cdot LinkDensity + \beta 2 \cdot NodeDensity + \beta 3 \cdot log(LotArea) + \beta 4 \\ &\cdot i.Style + \beta 5 \cdot i.Stories + \beta 6 \cdot YearBlt + \beta 7 \cdot YearBlt^2 + \beta 8 \\ &\cdot i.ExteriorFinish + \beta 9 \cdot i.Roof + \beta 10 \cdot i.Basement + \beta 11 \cdot i.Condition \\ &+ \beta 12 \cdot i.TotalRooms + \beta 13 \cdot i.Bedrooms + \beta 14 \cdot i.FullBaths + \beta 15 \\ &\cdot i.HalfBaths + \beta 16 \cdot i.Fireplaces + \beta 17 \cdot i.BsmtGarage + \epsilon \end{aligned}$$

In this equation, each term represents a different attribute that can relate to the fair market value of a property:

1. log(FairMarketTotal) represents the log transformed property fair market values, as listed in the Allegheny County dataset. These fair market values are adjusted for the dataset each year to ensure the comparability of the property appraisal values regardless of the latest transaction dates.
2. LinkDensity and NodeDensity are metrics of urban connectivity. They measure the density of ‘links’, which are the roads or paths that connect different places, and the ‘nodes’, which are the intersections or endpoints where these links meet or end. These variables are measured at the spatial scale of zip code areas and per square mile. Together, these counts also give us a number, a link/node ratio, that helps to understand how well-connected each zip code area is. This ratio is useful because it combines both link and node densities into a single metric, providing a broad measure of connectivity for the heterogeneity analyses. A lower ratio suggests that an area has more intersections and a more complex urban network, while a higher ratio indicates a simpler and easier-to-navigate rural network with more roads compared to intersections.
3. log(LotArea) adjusts for the log transformed size of the property, noting that larger properties are typically valued higher.
4. YearBlt represents the year the property was built, acknowledging that the age of a property associates with its market value, as newer properties or those with significant historical value may be valued differently. This relationship is modeled both linearly and quadratically to account for potential increases or decreases in value at different ages.
5. i.Style, i.ExteriorFinish, i.Roof, i.Basement, and i.Condition are categorical property level variables, where each category within these variables has been assigned a numeric code. The types of categories for each of the variables are listed in Appendix A. These variables

capture the relationship between architectural style, exterior finish, type of roof, basement characteristics, and the overall condition of the property and its value.

6. i.Stories, i.TotalRooms, i.Bedrooms, i.FullBaths, i.HalfBaths, and i.Fireplaces quantify the relationship between the number of stories, rooms, bedrooms, full and half bathrooms, and the presence of fireplaces, respectively, and property value, capturing the importance of these features in valuation.
7. i.BsmtGarage looks at the relationship between having a garage in the basement and the property's value.

Each of the coefficients ($\beta0,\beta1, ..., \beta17$) estimates the relationship between these variables and the logarithm of the fair market total value of a property, and ϵ represents the error term, capturing unobserved relationships. Among these variables, the variables of interest are LinkDensity and NodeDensity. The expected signs of the coefficients for LinkDensity and NodeDensity are positive, indicating that higher densities of links and nodes, which signify better connectivity, are associated with higher property appraisal values.

Additionally, heterogeneity analyses were conducted within different zip code areas representing varying levels of urban density within Allegheny County, determined by their respective link/node ratios. To assess how the relationship of connectivity differs between less densely connected more rural outer-city areas and more densely connected more urban inner-city areas, the dataset was split into two groups based on the median link/node ratio. Zip codes with a link/node ratio above the median were classified as less densely connected due to having more links relative to nodes, while those below the median were categorized as more densely connected. The study employed Chow F tests to assess the statistical significance of differences in the relationships between urban connectivity and property appraisal values between these two groups. This approach aims to provide a clearer understanding of how urban density relates to property fair market values across different neighborhood densities in the county. It is expected that the coefficients in more dense zip codes are less positive than in less dense zip codes.

3.2 Data Collection and Handling

The data for this research are from two main datasets, the Allegheny County Property Transactions Database and the National Neighborhood Data Archive on Urban Connectivity. The first dataset provides detailed information on property transactions within the county, encompassing a variety of variables that include estimated fair market value, year of construction, number of rooms, square footage, and other relevant attributes. The second dataset provides a clear picture of how easy it is to get around in each zip code area of Allegheny County by counting the links and nodes for each zip code area in Allegheny County.

The data management and cleaning are done by first combining the integration of the Allegheny County Property Transactions Database with the National Government Database on Urban Connectivity using ArcGIS and Stata. This integration (spatially) aligns property transaction data with urban connectivity metrics within each zip code area. To analyze connectivity, new variables, link and node density per zip code area, are calculated by dividing the number of links and nodes by the zip code area size in square miles. The link/node ratio variable is calculated by dividing the total links with the nodes for each zip code area. Once the datasets are merged and the variables are created, cleaning involves removing irrelevant variables, focusing the analysis on relevant variables, and excluding missing data. Unrealistic values, specifically extremely low property appraisal values (observations below 50000$ fair market value), are also eliminated after a manual inspection of the data to prevent skewed results and ensure the data accurately reflects typical market conditions.

Following the cleaning process, the data undergo transformation and normalization. The variables showing property prices and lot areas will be transformed to a log transformation to normalize their distribution, which is important for regression analysis as it stabilizes variance and makes the data more suitable for linear modeling. To enhance comparability in the regression model, the variable that contains the year when a property was built, was transformed by subtracting each value from 2022, making property age more simply quantified and allowing the variable to be squared.

In the regression analysis, it is also important to check the underlying assumptions to ensure the model's fit. The linearity of the relationship between the dependent and independent variables is examined by comparing observed values with predicted values to confirm that a linear relationship holds. Checking for homoscedasticity is also critical. This assumption requires that the variance of the residuals is constant across different values of the independent variables. Residual-versus-fitted plots are used to see if residuals spread evenly around the horizontal line without any discernible patterns. The assumption that errors are normally distributed is evaluated by looking at histograms of the residuals. Ideally, these histograms should closely follow a normal distribution curve, which is important for the validity of the model. The assumption of independent errors is already expected in the model. This is due to the nature of the data, as it is assumed that the data collection methodologies do not introduce dependency among observations. Finally, multicollinearity is considered by reviewing the Variance Inflation Factors (VIFs) for the independent variables. High VIFs would suggest that some variables are too closely related, which could cloud the regression estimates.

By carefully examining these assumptions, the analysis aims to ensure that it is conducted appropriately. However, it is important to note that in some cases, not all assumptions are perfectly met. This is also true for this regression model, where some assumptions are slightly violated. The consideration of these violations is necessary for accurately interpreting the results and understanding their implications.

Documentation of each step in the data handling and cleaning process as described in this section is also maintained through a detailed do-file. This documentation shows the steps taken to analyze the data, clear missing values, and other specifics of any data transformations, it is useful for replicability and transparency. It provides a clear guide to the methodologies used in preparing the data for analysis. All the Stata do-file code can be found in Appendix B, and the graph outputs resulting from these processes are included in Appendix C.

Ethical considerations include ensuring confidentiality, maintaining data integrity, and being transparent about methodologies and study limitations. Furthermore, data are securely stored on encrypted devices during the research period and will be permanently deleted within 1 year after the completion of the study to ensure data privacy. Appendix D contains more details about the upheld data management principles.

3.3 Descriptive Statistics & Visual Analysis

The following section delves into the descriptive statistics and visual analysis. Table 1 showcases the descriptive statistics for variables that are important for analyzing the relationship between urban connectivity and property appraisal values in Allegheny County. It provides information on the mean, standard deviation, minimum values and maximum values within the data.

Table 1: Descriptive Statistics of Key Variables

Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Fair Market Value	151486	230356	50100	7.23e+07
Node Density (per mi²)	.117	.0401	.0199	.228
Link Density (per mi²)	.412	.0268	.296	.486
Link Node Ratio	1.232704	.0802	.828	1.45
Lot Area	21565	319161	350	1.34e+08
Year Built	1948	27.2	1755	2022
Total Rooms	6.63	1.68	1	87
Bedrooms	3.12	.861	0	14
Full Bathrooms	1.48	.670	0	12
Half Bathrooms	.528	.577	0	9
Fireplaces	.445	.573	0	12
Garages	.807	.826	0	6
The dataset includes 127 unique zip code variable observations and 319,482 property-level observations				

Table 1 reveals the details about the properties in Allegheny County, such as fair market value, connectivity, and property features. The average property value is $151,486, but there is a wide range, with some properties valued as high as $72,300,000. This high maximum shows a few extremely valuable properties, mostly in the main downtown zip code area, that affect the average. Urban connectivity is measured by node density and link density, indicating how many intersections and roads are in each area, and shows a fairly consistent road network across the county. Property characteristics vary widely, with lot sizes ranging from small urban lots to very large rural properties. The ages of the properties also vary, with some constructed in the 1700s and others built as recently as 2022. Most homes have around six rooms, with typical numbers of bedrooms and bathrooms. Overall, the data shows a good mix of property fair market values and sizes, a consistent road network, and a variety of property ages and features.

Next to this, through the creation of Figures 3 through 6, this study also explored the visual relationship between urban connectivity and property appraisal values in Allegheny County. These maps aggregate infrastructure and property data at the zip code level to show the distribution of transportation links and nodes, their ratios, and average property fair market values in terms of area. Figures 3 and 4 reveal the densities of transportation infrastructure, with a focus on roads, bridges, and intersections. Figure 5 highlights the ratio of links to nodes, reflecting the transportation network's complexity and efficiency. Figure 6 presents the average property fair market values per zip code area. A pattern does seem to emerge from these visualizations: inner-city areas exhibit higher densities of transportation infrastructure and correspondingly higher property appraisal values, signifying a link between well-developed urban connectivity and real estate value. This trend is prevalent throughout most of Allegheny County, though some outer city areas present exceptions, suggesting the influence of additional factors on property appraisal values beyond just connectivity. The correlation is more clearly visualized using the scatter plots that can be seen in Figures 7, 8, and 9, where a fitted line clearly shows the relationship between the variables. The maps and scatter plots provide some insight for understanding how urban connectivity relates to property appraisal values, showing the significant variance between inner- and outer-city areas, with notable exceptions, and offering insights into the dynamics of the county's real estate market. Notably, there is an outlier in the scatter plot corresponding to the main downtown zip code area, which contains many high-rise buildings and is therefore worth significantly more.

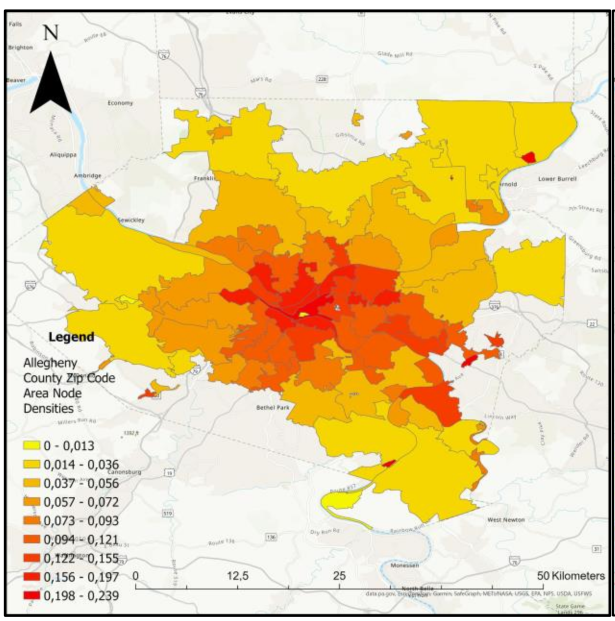


Figure 3: Node Densities per Zip Code Area

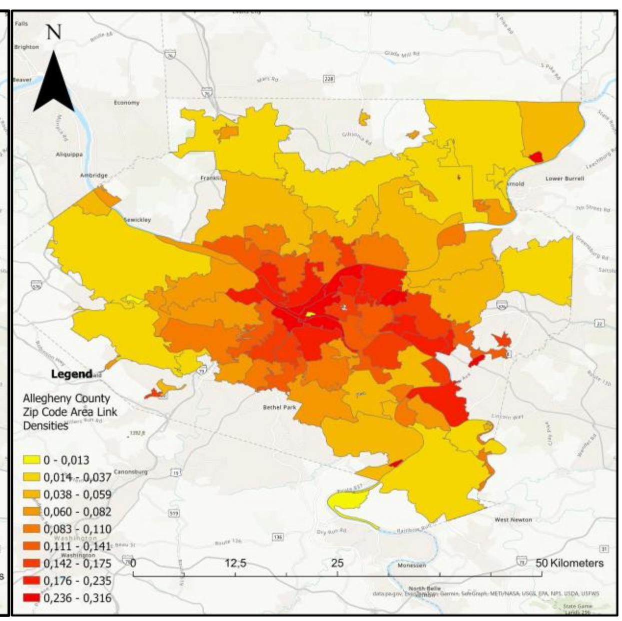
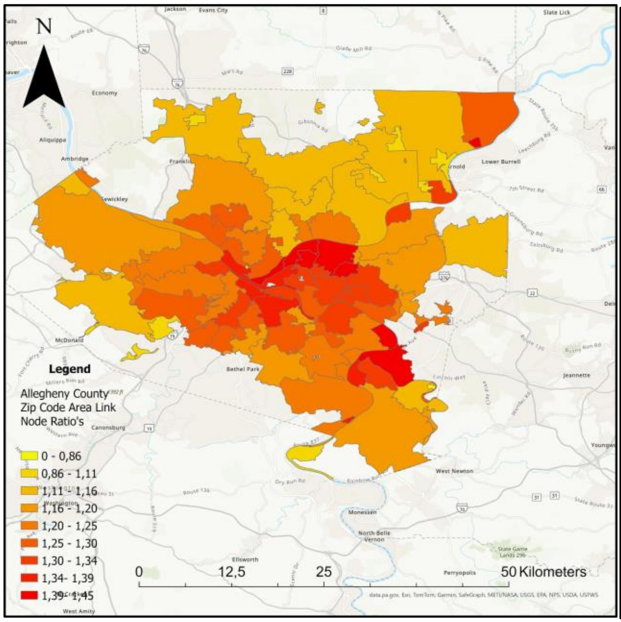


Figure 4: Link Densities per Zip Code Area



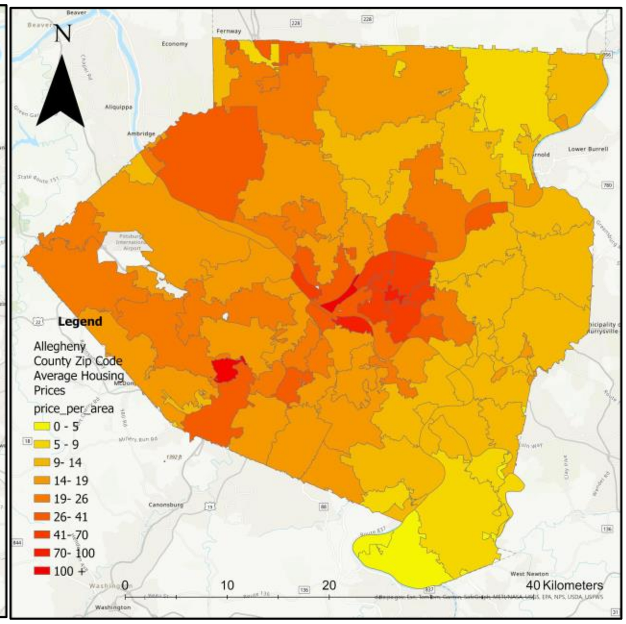
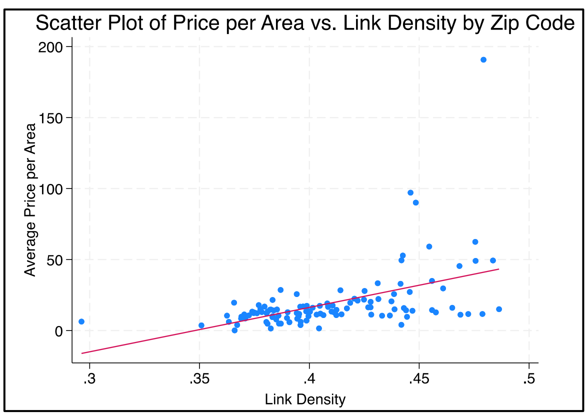


Figure 5: Link Node Ratio per Zip Code Area Figure 6: Average Appraisal Values per Zip Code Area



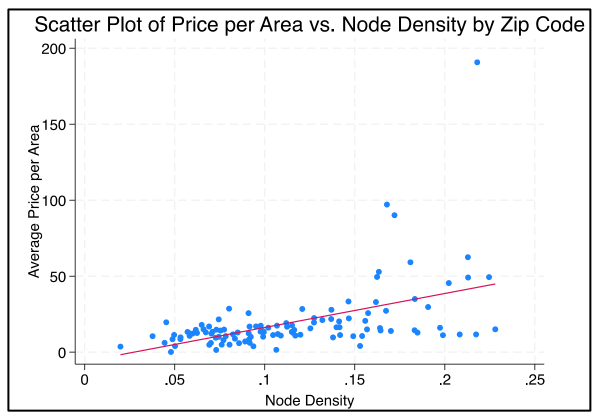


Figure 7: Average Appraisal Values per Link Densities Figure 8: Average Appraisal Values per Node Densities

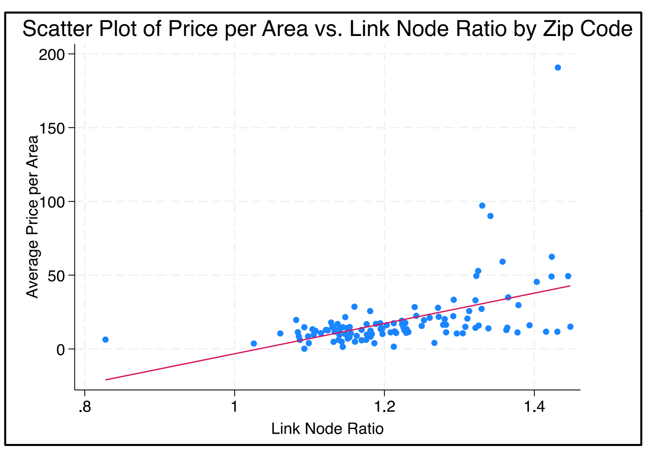


Figure 9: Average Property Fair Market Values per Link Node Ratios

4. Results & Discussion

4.1 Main Regression Results

Table 2 reports the regression analyses conducted to understand the determinants of property fair market values in Allegheny County. With a substantial sample size of 319,363 observations, the two full models show good explanatory power, as indicated by probabilities close to 0.0000, suggesting that the predictors significantly relate to the (logarithm of) fair market values of properties. This very low p-value indicates a statistically significant model overall, meaning that the likelihood of the results occurring by chance is extremely low. The models account for approximately 66% of the variance in property appraisal values (R-squared = 0.6571), which is comparable to other studies in the literature. The sizes of my coefficients vary from those in other research due to differing ways of measuring infrastructure and variations in study design, such as being conducted over time or in relation to single infrastructure projects. However, the explanatory power of the models in this study falls within the expected range of other research, with studies like Bujanda & Fullerton (2017) and Ghosh et al. (2021) reporting R-squared values in the range of 30% to 60% for similar, broader models, and You & Wong (2005) having an R-squared of even 85% in more specific contexts.

Table 2: Notable Regression Models and Results

Variable	Simple Model 1	Simple Model 2	Full Model 1	Full Model 2
Node Density (per mi²)	-	-.00417 (.0292)	-	.276* (.0216)
Link Density (per mi²)	.0188 (.0437)	-	.436* (.0323)	-
Log Lotarea	.245* (.00139)	.245* (.00139)	.136* (.00119)	.136* (.00119)
Year Built	-	-	-.0133* (.000167)	-.013320* (.000168)
Year Built ^ 2	-	-	.0000662* (1.10e-06)	.0000664* (1.10e-06)
Property Control Vars	No	No	Yes	Yes
Intercept	9.502* (.0269)	9.512* (.0146)	10.480* (.0318)	10.628* (.0274)
N	319,363	319,363	319,363	319,363
R-Squared	0.184	0.185	0.657	0.657
F Statistic	36079.73	36120.52	11334.52	113331.19
RSS	79759.8	79710.7	33529.0	33515.9
k	3	3	58	58
Dependent Variable is Log Fair Market Values | Robust Standard Errors in Parentheses | *p<0.001				

Looking at the specifics and starting with the control variables, the results align with expectations and the existing literature and confirm the theoretical underpinning of variation in consumer preferences and the importance of architectural design in real estate valuation (Thang, 2001; Margulis, 1988).

Of course, however, the study has a special focus on the role of urban connectivity. Specifically, connectivity as measured through the metrics link density and node density. Due to the presence of multicollinearity, as indicated by high variance inflation factors between link density and node density, separate models were used to analyze their individual relationships accurately.

The simple models indicate that the connectivity variables alone do not significantly influence property appraisal values. However, the inclusion of control variables in the full models reveals a different story. The first full model focuses on link density. The results support that an increase in link density is positively associated with property fair market values, as shown by a significant coefficient of 0.4362. To interpret this, consider the mean link density of 0.412 with a standard deviation of 0.0268. Since the dependent variable is the log-transformed fair market values, the coefficient can be interpreted as follows: a one standard deviation increase in link density (0.0268) results in approximately a 1.17% increase in property appraisal values. This finding highlights the importance of well-constructed road networks in boosting the attractiveness and value of properties. It shows that accessible, well-connected areas are preferred because they reduce travel times and costs when reaching services and amenities.

The second full model assesses the relationship between node density and property appraisal values. Here, a higher node density correlates positively with property fair market values, with a significant coefficient of 0.2761. Considering the mean node density of 0.117 with a standard deviation of 0.0401, a one standard deviation increase in node density (0.0401) results in approximately a 1.11% increase in property appraisal values. This supports the view that more intersections, which typically enhance connectivity within urban settings, contribute positively to property appraisal values. Intersections can be important for efficient urban travels, suggesting that properties in these areas offer better access to urban centers and transport hubs.

By interpreting these coefficients in the context of their means and standard deviations, it becomes clear that both link density and node density play significant roles in enhancing property fair market values through improved urban connectivity. These findings show the value of well-planned infrastructure in real estate development and urban planning. The results confirm the hypothesized preference for living in neighborhoods that offer convenient access to multiple locations, a finding that agrees with the literature about the critical role of accessibility and infrastructure in residential location decisions (Peris & Enault, 2023). The observation of the positive relationship of link and node densities on property appraisal values illustrates that areas with well-developed transportation infrastructure, which facilitates easy access to amenities, tend to be more desirable. This is further supported by Haugen (2011), who noted the value buyers place on proximity to work, services, and social activities, and by the mentioned practical studies by Cohen & Schaffner (2021) and Yiu & Wong (2005), which show the significant influence of infrastructure development on housing prices. These studies offer evidence that infrastructure improvements, like new highways or anticipated transportation projects (e.g., tunnels), can lead to significant increases in property values.

Therefore, the analysis, using the hedonic pricing model, both supports the relationship of general factors such as property size and condition to values but also shows the complex and positive relationship of urban connectivity. When it comes to the first hypothesis, which stated that specific attributes of urban connectivity, the number of links and nodes, would significantly and positively relate to property fair market values, the findings show that this hypothesis cannot be rejected. The significant coefficients for link density (0.4362) and node density (0.2761) show the substantial positive association of urban connectivity features with property appraisal values.

4.2 Heterogeneity Analyses Results

Table 3: Notable Regression Models and Results of Less and More Dense Areas in Allegheny County

Variable	Less Dense Model 1	Less Dense Model 2	More Dense Model 1	More Dense Model 2
Node Density (per mi²)	-	.282* (.0374)	-	-3.731* (.0525)
Link Density (per mi²)	.434* (.0560)	-	-5.070* (.0929)	-
Property Control Vars	Yes	Yes	Yes	Yes
Intercept	10.488* (.0488)	10.635* (.0402)	-13.422* (.383)	-15.004* (.377)
N	154,728	154,728	164,754	164,754
R-Squared	0.554	0.554	0.7504	0.7513
F Statistic	3560.6	3560.5	9341.9	9382.8
RSS	19938.1	19938.5	12372.2	12314.1
k	58	58	58	58
Dependent Variable is Log Fair Market Values | Robust Standard Errors in Parentheses | *p<0.001				

In the analysis of how urban connectivity relates to property fair market values in Allegheny County, another aspect of the study involves examining whether the relationship differs between areas with varying degrees of urban density. For this purpose, the dataset is divided into two groups based on the median link/node ratio, which serves as a representation for urban density. Properties in zip codes with a link/node ratio above the median were classified as "Less Dense”, showing less urban connectivity relative to node density. Next to this, those with a ratio below the median were classified as "More Dense”, indicating higher urban connectivity. The regression results that followed from this are shown in Table 3, but the more detailed regression results and specific Chow F Test statistic calculations can be found in Appendix F.

To determine if the relationships between urban connectivity and property appraisal values significantly differ between these two groups, the Chow F Test is used. This statistical test compares two linear regression models, each representing one of the density groups, to a combined model that includes all data. The Chow F Test calculates a statistic by comparing the sum of the residual squares from the separate models for each group to the residual sum of squares from a single model that includes all data. The formula for the Chow F statistic is:

$$F = \frac{(\mathrm{RSSC} - (\mathrm{RSS1} + \mathrm{RSS2}))/\mathrm{k}}{(\mathrm{RSS1} + \mathrm{RSS2})/(\mathrm{n1} + \mathrm{n2} - 2\mathrm{k})}$$

Here, RSSC is the residual sum of squares from the combined model, RSS1 and RSS2 are the residual sums of squares from the models for the "More Dense" and "Less Dense" groups, respectively. n1 and n2 are the numbers of observations in each group, and k represents the number of parameters (including the intercept) estimated in each regression.

The high F-value of 207.7 resulting from the aforementioned calculation for the Less Dense models shows that the models are significantly better fits when ran separately rather than as a single combined model, suggesting that link density relates to property appraisal values differently across urban densities. Similarly, for the model including the node density variable, the high F-value of 215.7 supports the hypothesis that the relationship between urban connectivity, as indicated by node density, also vary significantly between more densely and less densely connected areas. Based on these results, the separate models for the "More Dense" and "Less Dense" areas provide a significantly better fit to the data than a single combined model. This result also aligns with the literature, as Mathur (2008) pointed out that the relationship between property values depends on the existing urban infrastructure, which can highly vary between neighborhoods.

As for the coefficients, in less dense areas, the coefficient for link density is 0.434, which is positive and statistically significant. Given the mean link density of 0.412 and a standard deviation of 0.0268, a one standard deviation increase in link density results in about a 1.16% increase in property appraisal values. Similarly, the coefficient for node density in the less dense model is 0.282, which is also positive and statistically significant. With a mean node density of 0.117 and a standard deviation of 0.0401, a one standard deviation increase in node density results in approximately a 1.13% increase in property appraisal values. These positive relationships suggest that in less densely connected areas, improvements in both road networks and intersections significantly enhance property fair market values. As the bid rent theory suggests, better connectivity reduces transportation costs and increases accessibility, making these areas more attractive to potential buyers (Alonso, 1964). Improved accessibility leads to higher demand for properties, which leads to higher property values. Therefore, investments in road and intersection infrastructure in these less dense areas can have substantial positive effects on real estate markets.

In contrast, the more dense model reveals a different story. The coefficient for link density in more dense areas is -5.070, which is negative and statistically significant. For a mean link density of 0.412 and a standard deviation of 0.0268, a one standard deviation increase in link density results in a 13.6% decrease in property appraisal values. Similarly, the coefficient for node density in the more dense model is -3.731, also negative and statistically significant. Given a mean node density of 0.117 and a standard deviation of 0.0401, a one standard deviation increase in node density results in a 14.9% decrease in property appraisal values. These major negative relationships indicate that in more densely connected areas, increases in both link and node density lead to decreased property fair market values. According to the bid rent theory, as transportation costs decrease due to improved connectivity, the relative advantage of being closer to the city center becomes lower (Alonso, 1964). This flattening of the bid rent curve means that properties in more densely connected inner areas become less attractive compared to those in outer areas. Excessive road and intersection density in already densely connected urban settings might lead to congestion, noise, and other negative externalities as well, which outweigh the benefits of improved accessibility, resulting in lower property values.

These heterogeneity analyses show the complex and context-dependent nature of the relationship between urban connectivity and property appraisal values. In less densely connected areas, improvements in both link and node density significantly enhance property appraisal values, supporting the idea that better connectivity reduces transportation costs and increases desirability. On the other hand, in more densely connected areas, excessive increases in link and node density lead to significant decreases in property appraisal values, showing the negative externalities of over-congestion and decrease in premiums willing to be paid compared towards the outer areas. These findings are consistent with the theoretical insights provided by the bid rent theory, illustrating how accessibility influences real estate values across different urban settings. Therefore, we cannot reject the hypothesis that the relationship between urban connectivity and property fair market values is significant and greater in less densely connected inner-city areas compared to more densely connected outer-city areas. These findings show the differing associations of urban infrastructure with real estate values, emphasizing the need for more specific urban planning and policy-making that looks at the contextual characteristics and needs of different neighborhoods.

5. Conclusion

This thesis has explored the relationship between urban connectivity and property appraisal values in Allegheny County, providing insights into how transportation infrastructure relates to the real estate market. The findings align with the existing literature, which shows the significance of well-developed transport networks in increasing property values. However, this study also brings new perspectives, particularly in the analysis of how different features of connectivity, such as the density of roads and intersections, can relate to property appraisal values.

The results presented in this thesis support the significant positive relationship between urban connectivity and property appraisal values. Well-planned intersections and an efficient road network, which enhance accessibility to various amenities and locations, are shown to significantly boost property appraisal values. This finding is consistent with theories in urban economics that show the importance of accessibility in real estate valuation. Another important insight from this study is the variation in the relationship between urban connectivity and property fair market values between more densely connected inner-city areas and less densely connected outer-city areas. In less densely connected areas, increased link and node densities are associated with higher property appraisal values, indicating the benefits of improved accessibility. However, in more densely connected areas, additional links and nodes have a negative relationship with property appraisal values. Reduced transportation costs make outer areas more attractive, diminishing the premium on inner city properties. Next to this, in densely connected areas, excessive infrastructure can lead to a reduced premium and negative externalities that seem to outweigh the benefits of increased connectivity.

For policymakers and urban planners, these insights show the importance of focusing on well-designed and well-maintained transportation networks that improve connectivity in the right places. The differences between densely connected inner-city areas and less connected outer-city areas show the need for strategic infrastructure development, meaning that the quality and strategic placement of infrastructure improvements are more important than just increasing the quantity of roads wherever possible.

When it comes to the limitations of the results, it is important to recognize the different scale levels at which the variables are measured. The independent variables regarding urban connectivity, density of links and nodes, and the link/node ratios are aggregated at the zip code level, which is broader than the property-level scale of the other variables. When interpreting the consistency and efficiency of the estimates, this scale difference should be considered during the analysis. Additionally, potential omitted variable bias should be noted, as many factors such as local economic conditions, school quality, and crime rates, which can also influence property values, are not included in the model. The link/node metrics, while useful, might oversimplify the complexity of urban infrastructure by not capturing other aspects of connectivity like public transportation and pedestrian or bike infrastructure. Lastly, the study uses cross-sectional data, which limits the ability to understand changes over time. Longitudinal data could provide a better view of how changes in connectivity can impact property appraisal values.

For future research, investigation into the specific types of infrastructure that most significantly relate to property values could be useful, together with in which locations these interventions could be most positive. This could include a focus on more specific conventional infrastructure features, such as tunnels, highways and public transportation, or more progressive concepts including sustainable infrastructure like green parks or energy-efficient public transport systems. This can help understand how these elements contribute to property valuations in urban areas.

In summary, this thesis not only researches the diverse factors influencing property appraisal values in Allegheny County, including the important role of urban connectivity, but also contributes to a better understanding of how these features interact within the broader real estate market. The insights obtained from this study can offer a basis for informed decision-making and strategic planning for future urban development. It can be important for everyone from homeowners and buyers to policymakers and urban planners to be more informed and effective in creating urban development strategies.

References

Adair, A. (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. Housing Studies, 15(5), 699-716.

Alonso, W. (1964). Location and Land Use. Toward a General Theory of Land Rent. Cambridge: Harward University Press.

Bujanda, A. & Fullerton, T.M. (2017). Impacts of transportation infrastructure on single-family property values. *Applied Economics,* 49, 5183-5199.

Cohen, J.P., & Schaffner, S. (2021). New German Highway Infrastructure and the Impacts on Residential Real Estate Prices. *ERN: Urban Markets (Topic)*, 1-32. *Transportation Research Part A: Policy and Practice*, 153, 218-234.

Demİrdağ, S. (2023). Examining variables affecting housing prices within the scope of transportation infrastructure. Journal of Urban Planning and Development, 149(2), 04022045.

Deng, X., Nanda, A., & Ong, S.E. (2019). Does infrastructure spending lead to price effects in the property market? Evidence from major cities across India. *Regional Studies,* 53, 1747-1760.

Dorantes, A. (2011). Analysis of house prices to assess economic impacts of new public transport infrastructure. Journal of Transport and Land Use, 4(3), 45-58.

Ghosh, C., Panchapagesan, V., & Venkataraman, M. (2021). On the Impact of Infrastructure Improvement on Real Estate Property Values: Evidence from a Quasi-natural Experiment in an Emerging Market. *The Journal of Real Estate Finance and Economics*, 1-35.

Haugen, K. (2011). The Advantage of 'Near': Which Accessibilities Matter to Whom? *European Journal of Transport and Infrastructure Research*, 11(4), 368-388.

Hoogendoorn, S. (2019). House prices and accessibility: Evidence from a quasi-experiment in transport infrastructure. Regional Science and Urban Economics, 77, 1-12.

Evans, A. W. (2008). *Economics, Real Estate and the Supply of Land*. Hoboken: Wiley-Blackwell.

Margulis, H.L. (1988). Homebuyer choices and search behaviour in a distressed urban setting. *Housing Studies,* 3, 112-133.

Matur, S. (2008). Impact of Transportation and Other Jurisdictional-Level Infrastructure and Services on Housing Prices. *Journal of Urban Planning and Development,* 134(1), 32-41.

Mikelbank, B. A. (2000). Spatial analysis of the relationship between transportation infrastructure and metropolitan housing values: Integrating traditional, spatial, and geostatistical approaches. Journal of Urban Economics, 48(1), 1-21.

Mischke, J., Luby, R., Vickery, B., Woetzel, J., White, O., Sanghvi, A., Rhee, J., Fu, A., Palter, R., Dua, A. & Smit, S. (2023). *Empty spaces and hybrid places: The pandemic's lasting impact on real estate.* McKinsey Global Institute.

Peris, A., & Enault, L. C. (2023). Proximity or opportunity? Spatial and market determinants of private individuals' buy-to-let investments. *Environment and Planning B: Urban Analytics and City Science*, 1-23.

Pressley, S. (2023). Infrastructure Influence on Property Markets. *Propertyology,* 06-10-2023.

Shaw, V. (2022). Prime property prices in urban locations 'rising faster than in rural areas'. *Independent,* 11-07-2022.

Shin, H. (2007). Effects of transportation accessibility on residential property values. Transportation Research Part A: Policy and Practice, 41(6), 549-560.

Thang, D.C., & Eng, O.S. (2001). Mobility propensity of Public Housing Residents. *8th European Real Estate Society Conference.*

Wikimedia Foundation (2023). *History of Pittsburgh*. Retrieved on 23-10-2023 from https://en.wikipedia.org/wiki/History_of_Pittsburgh.

Wikimedia Foundation (2023). *North Shore Connector*. Retrieved on 23-10-2023 from https://en.wikipedia.org/wiki/North_Shore_Connector.

Yang, J. (2020). Accessibility to transit, by transit, and property prices: Spatially varying relationships. Journal of Transport Geography, 82, 102-110.

You, C.Y. & Wong, S.K. (2005). The Effects of Expected Transport Improvements on Housing Prices. *Urban Studies,* 42(1), 113-125.

5. Appendices

5.1 Appendix A: Assignment of Numeric Codes to Categorical Variables

Variable	Numeric Code	Description	Variable	Numeric Code	Description
Style	1	Ranch	Exterior Finish	1	Frame
	2	Split Level		2	Brick
	3	Bi-Level		3	Stone
	4	Colonial		4	Stucco
	5	Cape Cod		5	Concrete Block
	6	Conventional		6	Masonry Frame
	7	Contemporary		7	Concrete
	8	Condo		8	Log
	9	Townhouse	Roof	1	Shingle
	10	Modular Home		2	Slate
	11	Row End		3	Metal
	12	Row Interior		4	Roll
	13	Multi-Family		5	Tile
	14	Victorian		6	Rubber
	15	Other	Basement	1	None
	16	Old Style		2	Slab/Piers
	17	Log Cabin		3	Crawl
	18	Bungalow		4	Part
	19	Tudor		5	Full
	20	Semi Detached	Condition	1	Excellent
	24	Condo End		2	Good
	25	Condo Int		3	Average
				4	Fair
				5	Poor
				6	Unsound
				7	Very Good
				8	Very Poor

5.2 Appendix B: Do-File Steps

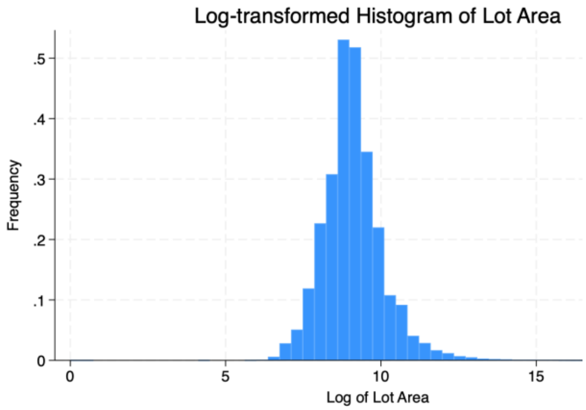
```
*** Importing Dataset ***
import delimited "FullyMergedFile.csv", clear

*** Data Cleaning ***
* Renaming and Adjusting Variables
rename notes notes_var
rename gamma linkdensity
rename alpha nodedensity
drop if style == "M1" | style == "M2"
destring style, replace
* Dropping Unnecessary Variables
drop objectid mapblocklo municode calcacreag pseudono comments globalid shape_leng
shape_area propertyfraction changenoticeaddress1 changenoticeaddress3
changenoticeaddress4 countyexemptbldg alt_id taxyear asofdate _merge intdensity
strnetdensity avgblocklength medblocklength zcta_area_sqmiles sum_strintlen n_blocks
taxcode taxdesc ownerdesc usecode homesteadflag farmsteadflag cleangreen abatementflag
recorddate saledate saleprice salecode saledesc deedbook deedpage prevsaledate
prevsaleprice prevsaledate2 prevsaleprice2 countybuilding countyland countytotal localbuilding
localland localtotal grade gradedesc cdu cdudesc finishedlivingarea cardnumber notes_var

* Handling Missing Values
* Assuming missing values are coded as ".", NA, or blank.
drop if oid_ == . | propertyhousenum == . | propertyzip == . | lotarea == . | fairmarketbuilding == .
| fairmarketland == . | fairmarkettotal == . | stories == . | yearblt == . | exteriorfinish == . | roof ==
. | basement == . | condition == . | totalrooms == . | bedrooms == . | fullbaths == . | halfbaths == .
| fireplaces == . | bsmtgarage == . | n_streets == . | n_nodes == . | n_realnodes == . |
linknoderatio == . | connoderatio == . | blockdensity == . | parid == "" | neighcode == "" | usedesc
== "" | style == . | styledesc == "" | extfinish_desc == "" | roofdesc == "" | basementdesc == "" |
conditiondesc == "" | heatingcooling == "" | heatingcoolingdesc == "" | nodedensity == . |
linkdensity == .

* Checking for and Manually Removing Unrealistic Values
sum fairmarkettotal nodedensity linkdensity linknoderatio lotarea yearblt totalrooms bedrooms
fullbaths halfbaths fireplaces bsmtgarage
drop if fairmarkettotal < 50000

* Generate (log-)transformed variables for better regression results
gen log_lotarea = log(lotarea)
gen log_fairmarkettotal = log(fairmarkettotal)
gen yearbltedited = 2022 - yearblt
gen yearbltedited2 = yearbltedited^2
```

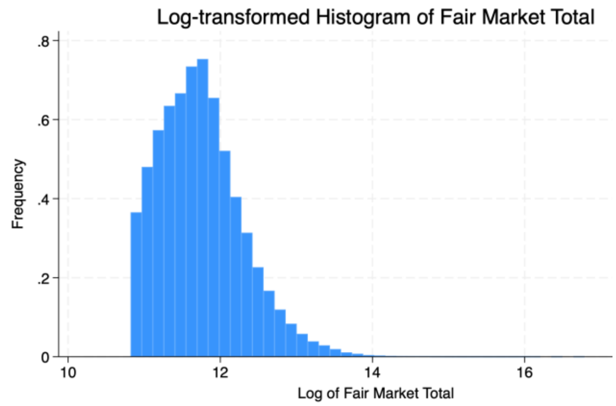
```
histogram log_lotarea, title("Log-transformed Histogram of Lot Area") xtitle("Log of Lot Area")
ytitle("Frequency") bin(50) graphregion(color(white)) plotregion(color(white))
histogram log_fairmarkettotal, title("Log-transformed Histogram of Fair Market Total") xtitle("Log
of Fair Market Total") ytitle("Frequency") bin(50) graphregion(color(white))
plotregion(color(white))

* Save cleaned and transformed dataset and import if needed
save "cleaned_and_transformed_data.csv", replace

*** Regression Analyses ***
* Multiple Linear Regression
* Model 1
reg log_fairmarkettotal linkdensity log_lotarea, robust

* Model 2
reg log_fairmarkettotal nodedensity log_lotarea, robust

* Model 3
reg log_fairmarkettotal linkdensity log_lotarea yearbltedited yearbltedited2 i.style i.exteriorfinish
i.roof i.basement i.condition totalrooms bedrooms fullbaths halfbaths fireplaces bsmtgarage,
robust

* Model 4
reg log_fairmarkettotal nodedensity log_lotarea yearbltedited yearbltedited2 i.style
i.exteriorfinish i.roof i.basement i.condition totalrooms bedrooms fullbaths halfbaths fireplaces
bsmtgarage, robust

*** Assumption Checking ***
* Checking Linearity
* Generate Scatter Plots with Fitted Lines
gen price_per_area = fairmarkettotal / lotarea
egen avg_price_per_area = mean(price_per_area), by(propertyzip)
duplicates drop propertyzip, force
twoway (scatter avg_price_per_area linkdensity) (lfit avg_price_per_area linkdensity),
xtitle("Link Density") ytitle("Average Price per Area") title("Scatter Plot of Price per Area vs. Link
Density by Zip Code")
twoway (scatter avg_price_per_area nodedensity) (lfit avg_price_per_area nodedensity),
xtitle("Node Density") ytitle("Average Price per Area") title("Scatter Plot of Price per Area vs.
Node Density by Zip Code")
twoway (scatter avg_price_per_area linknoderatio) (lfit avg_price_per_area linknoderatio),
xtitle("Link Node Ratio") ytitle("Average Price per Area") title("Scatter Plot of Price per Area vs.
Link Node Ratio by Zip Code")

* Comparing Observed with Predicted Values
```

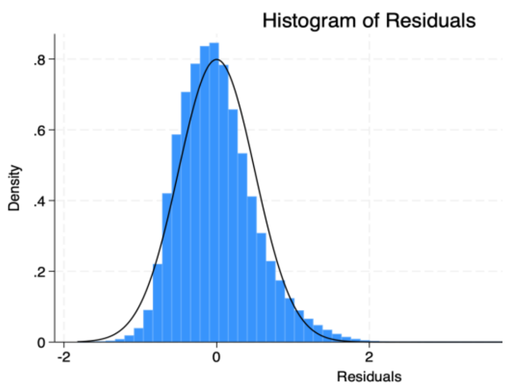
```
predict yhat
scatter log_fairmarkettotal yhat || lfit log_fairmarkettotal yhat

* Checking for Homoscedasticity
rvfplot, yline(0)

* Plot residuals to check for homoscedasticity
predict residuals, residuals
predict predicted, xb
scatter residuals predicted, title("Residuals vs. Predicted Values") xtitle("Predicted Values")
ytitle("Residuals")

* Histogram of residuals
histogram residuals, title("Histogram of Residuals") normal

* Omitted Variable Bias
ovtest

* Q-Q plot of residuals
qnorm residuals, title("Q-Q Plot of Residuals")

* Check for multicollinearity
vif

*** Chow F Test ***
* Calculate Median
su linknoderatio, detail
gen median_indicator = linknoderatio > r(p50)

* Split Group
gen group = "Less Dense" if median_indicator == 1
replace group = "More Dense" if median_indicator == 0

* Run Chow F Regression Models
reg log_fairmarkettotal linkdensity log_lotarea i.style i.exteriorfinish yearbltedited yearbltedited2
i.roof i.basement i.condition totalrooms bedrooms fullbaths halfbaths fireplaces bsmtgarage if
group == "Less Dense"

reg log_fairmarkettotal nodedensity log_lotarea i.style i.exteriorfinish yearbltedited
yearbltedited2 i.roof i.basement i.condition totalrooms bedrooms fullbaths halfbaths fireplaces
bsmtgarage if group == "Less Dense"
```

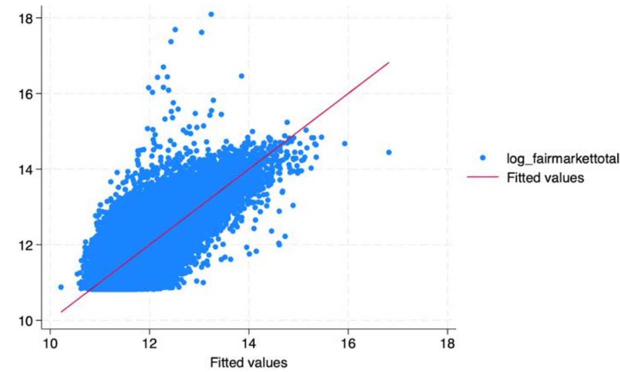
```
reg log_fairmarkettotal linkdensity log_lotarea i.style i.exteriorfinish yearblt yearbltedited2 i.roof
i.basement i.condition totalrooms bedrooms fullbaths halfbaths fireplaces bsmtgarage if group
== "More Dense"

reg log_fairmarkettotal nodedensity log_lotarea i.style i.exteriorfinish yearblt yearbltedited2 i.roof
i.basement i.condition totalrooms bedrooms fullbaths halfbaths fireplaces bsmtgarage if group
== "More Dense"

*** Creating GIS Maps ***
* Link / Node Density Maps
* These maps are created using ArcGIS and its calculator function
* It divides the Links and Nodes by the Zip Code Area sizes to give densities

* Real Estate Value Map
collapse (mean) price_per_area, by(propertyzip)
sum propertyzip price_per_area
export delimited using average_price_per_m2_by_zip.csv, replace
```

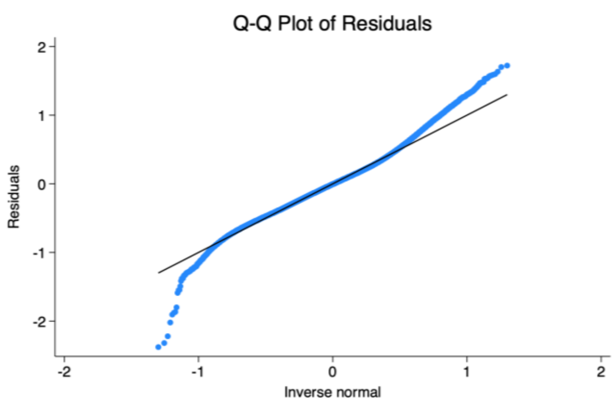
5.3 Appendix C: Stata Graph Outputs

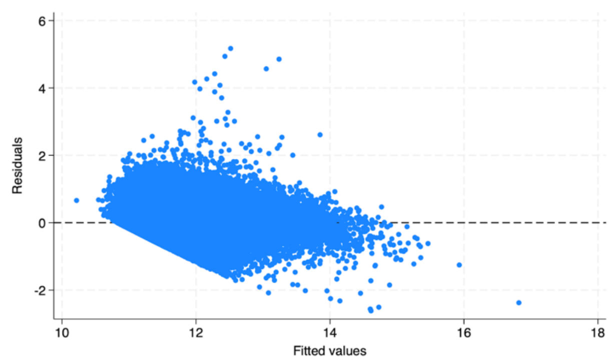












Variable	VIF	1/VIF
nodedensity	**1.70**	**0.587626**
log_lotarea	**1.87**	**0.534324**
yearbltedi~d	**29.84**	**0.033512**
yearbltedi~2	**25.68**	**0.038946**
style		
2	**1.17**	**0.851333**
3	**1.32**	**0.756815**
4	**2.16**	**0.463344**
5	**1.44**	**0.693086**
6	**1.00**	**0.997979**
7	**1.08**	**0.921969**
8	**1.00**	**0.999603**
9	**1.57**	**0.635149**
10	**1.01**	**0.988523**
11	**1.17**	**0.851825**
12	**1.26**	**0.794972**
13	**2.29**	**0.436385**
14	**1.04**	**0.965881**
15	**1.00**	**0.995930**
16	**3.26**	**0.307151**
17	**1.90**	**0.527149**
18	**1.15**	**0.868445**
19	**1.04**	**0.961808**
20	**1.19**	**0.836854**
24	**1.00**	**0.999708**
25	**1.00**	**0.999791**
exteriorfi~h		
2	**1.71**	**0.585865**
3	**1.13**	**0.885096**
4	**1.02**	**0.977691**
5	**1.01**	**0.988305**
6	**1.45**	**0.691627**
7	**1.00**	**0.997592**
8	**1.91**	**0.524420**
roof		
2	**1.16**	**0.860477**
3	**1.00**	**0.996017**
4	**1.17**	**0.851389**
5	**1.02**	**0.984692**
6	**1.00**	**0.997340**
basement		
2	**1.24**	**0.805896**
3	**1.29**	**0.775230**
4	**4.53**	**0.220950**
5	**5.06**	**0.197808**
condition		
2	**69.14**	**0.014463**
3	**124.91**	**0.008006**
4	**61.96**	**0.016139**
5	**5.71**	**0.175160**
6	**1.16**	**0.864692**
7	**8.05**	**0.124220**
8	**1.53**	**0.655434**
totalrooms	**3.49**	**0.286821**
bedrooms	**2.70**	**0.369911**
fullbaths	**1.96**	**0.511203**
halfbaths	**1.40**	**0.712437**
fireplaces	**1.31**	**0.761631**
bsmtgarage	**1.67**	**0.598554**
Mean VIF	**7.40**	

5.4 Appendix E: Data Management Principles

1. General	
1.1 Name & title of thesis	Evaluating Urban Connectivity and its Association with Property Appraisal Values in Allegheny County

2 Data collection – the creation of data	
2.1. Which data formats or which sources are used in the project? For example: - theoretical research, using literature and publicly available resources - Survey Data - Interviews	Theoretical Research (Academic Literature and Papers) and Two Publicly Available Datasets.
2.2 Methods of data collection What method(s) do you use for the collection of data. (Tick all boxes that apply)	☐ Structured individual interviews ☐ Semi-structured individual interviews ☐ Structured group interviews ☐ Semi-structured group interviews ☐ Observations ☐ Survey(s) ☐ Experiment(s) in real life (interventions) ☒ Secondary analyses on existing data sets ☐ Public sources (e.g. University Library) ☐ Other (explain):
2.3. (If applicable): if you have selected 'Secondary analyses on existing datasets': who provides the data set?	☐ Data is supplied by the University of Groningen. ☒ Data have been supplied by an external party. (The Allegheny County Property Transactions Database and the National Neighborhood Data Archive on Urban Connectivity).

3 Storage, Sharing and Archiving	
3.1 Where will the (raw) data be stored *during* research? If you want to store research data, it is good practice to ask yourself some questions:	☐ X-drive of UG network ☐ Y-drive of UG network ☐ (Shared) UG Google Drive ☐ Unishare

• How big is my dataset at the end of my research? • Do I want to collaborate on the data? • How confidential is my data? • How do I make sure I do not lose my data? Need more information? Take a look at the site of the Digital Competence Centre (DCC)) Feel free to contact the DCC for questions: dcc@rug.nl	☒ Personal laptop or computer ☐ External devices (USB, harddisk, NAS) ☐ Other (explain):
3.2 Where are you planning to store / archive the data after you have finished your research? Please explain where and for how long. Also explain who has access to these data NB do not use a personal UG network or google drive for archiving data!	☐ X-drive of UG network ☐ Y-drive of UG network ☐ (Shared) UG Google Drive ☐ Unishare ☐ In a repository (i.e. DataverseNL) ☒ Other (explain): On my personal laptop and computer The retention period will be 1< years. The data is publicly available.
3.3 Sharing of data With whom will you be sharing data during your research?	☐ University of Groningen ☐ Universities or other parties in Europe ☐ Universities or other parties outside Europe ☒ I will not be sharing data (it is not needed as it is public)

4. Personal data	
4.1 Collecting personal data Will you be collecting personal data? If you are conducting research with personal data you have to comply to the General Data Privacy Regulation (GDPR). Please fill in the questions found in the appendix 3 on personal data.	No

5 – Final comments	
Do you have any other information about the research data that was not addressed in this template that you think is useful to mention?	No

5.5 Appendix E: Stata Full Regression Models

Source	SS	df	MS			
Model	18015.0318	2	9007.51591	Number of obs	=	319,482
Residual	79759.8033	319,479	.249655856	F(2, 319479)	=	36079.73
				Prob > F	=	0.0000
Total	97774.8351	319,481	.306042723	R-squared	=	0.1843
				Adj R-squared	=	0.1842
				Root MSE	=	.49966

log_fairma~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
linkdensity	.018818	.0384433	0.49	0.624	-.0565298	.0941657
log_lotarea	.2453245	.0010641	230.54	0.000	.2432388	.2474101
_cons	9.501949	.0224306	423.62	0.000	9.457985	9.545912

Source	SS	df	MS			
Model	18031.0085	2	9015.50423	Number of obs	=	319,363
Residual	79710.6861	319,360	.249595084	F(2, 319360)	=	36120.52
				Prob > F	=	0.0000
Total	97741.6946	319,362	.306052989	R-squared	=	0.1845
				Adj R-squared	=	0.1845
				Root MSE	=	.49959

log_fairma~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
nodedensity	-.0041744	.0257263	-0.16	0.871	-.0545971	.0462483
log_lotarea	.2451261	.0010644	230.29	0.000	.2430399	.2472123
_cons	9.512194	.0115719	822.01	0.000	9.489514	9.534875

Source	SS	df	MS			
Model	64245.8723	54	1189.73838	Number of obs	=	319,482
Residual	33528.9628	319,427	.104965963	F(54, 319427)	=	11334.52
				Prob > F	=	0.0000
Total	97774.8351	319,481	.306042723	R-squared	=	0.6571
				Adj R-squared	=	0.6570
				Root MSE	=	.32398

log_fairmark~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
linkdensity	.4361939	.0278484	15.66	0.000	.3816119	.4907759
log_lotarea	.1363599	.0008091	168.53	0.000	.134774	.1379457

style						
2	.0505753	.0034337	14.73	0.000	.0438454	.0573052
3	.0282945	.0029831	9.48	0.000	.0224477	.0341414
4	.0813474	.0019971	40.73	0.000	.0774331	.0852617
5	-.0265769	.0022059	-12.05	0.000	-.0309004	-.0222534
6	.2718113	.0372056	7.31	0.000	.1988895	.3447332
7	.1945922	.006989	27.84	0.000	.1808939	.2082904
8	.2594366	.083671	3.10	0.002	.0954438	.4234295
9	.1257066	.004341	28.96	0.000	.1171984	.1342147
10	-.3632411	.0284763	-12.76	0.000	-.4190538	-.3074284
11	.2478082	.0076312	32.47	0.000	.2328513	.2627652
12	.2774149	.0069942	39.66	0.000	.2637065	.2911234
13	-.1347128	.0039721	-33.91	0.000	-.1424981	-.1269276
14	.2915904	.0144404	20.19	0.000	.2632877	.3198932
15	.2914531	.0765219	3.81	0.000	.1414723	.4414338
16	.1073534	.0024616	43.61	0.000	.1025287	.1121781
17	-.0238076	.0352863	-0.67	0.500	-.0929678	.0453525
18	-.0409555	.0048477	-8.45	0.000	-.0504569	-.0314541
19	.2037853	.0104787	19.45	0.000	.1832472	.2243233
20	.1283661	.0049644	25.86	0.000	.118636	.1380963
24	.3683982	.1870807	1.97	0.049	.0017255	.735071
25	.3927563	.187073	2.10	0.036	.0260986	.759414
exteriorfinish						
2	.0935513	.0015064	62.10	0.000	.0905989	.0965037
3	.1490283	.0050737	29.37	0.000	.1390839	.1589727
4	.1504646	.0081291	18.51	0.000	.1345318	.1663973
5	.033926	.0131334	2.58	0.010	.008185	.0596671
6	.0867988	.00201	43.18	0.000	.0828594	.0907383
7	.2707806	.0263166	10.29	0.000	.2192009	.3223604
8	-.0150524	.0285935	-0.53	0.599	-.0710948	.04099
yearbltedited	-.0133057	.0001151	-115.60	0.000	-.0135313	-.0130801
yearbltedited2	.0000662	6.50e-07	101.80	0.000	.0000649	.0000675
roof						
2	.1988642	.0027633	71.97	0.000	.1934481	.2042802
3	.0807621	.0196201	4.12	0.000	.0423073	.1192169
4	.066336	.0044911	14.77	0.000	.0575336	.0751384
5	.2952281	.0085907	34.37	0.000	.2783906	.3120655
6	.0914393	.0332893	2.75	0.006	.0261931	.1566854
basement						
2	.0852683	.0130834	6.52	0.000	.0596252	.1109114
3	-.0029856	.0119447	-0.25	0.803	-.026397	.0204257
4	.0440834	.0064245	6.86	0.000	.0314916	.0566752
5	.0132935	.0057399	2.32	0.021	.0020435	.0245436

condition						
2	-.316312	.0159994	-19.77	0.000	-.3476704	-.2849537
3	-.5482606	.0159472	-34.38	0.000	-.5795166	-.5170045
4	-.7310327	.016071	-45.49	0.000	-.7625314	-.6995339
5	-.8138196	.0175602	-46.34	0.000	-.8482371	-.7794022
6	-.7736892	.0435559	-17.76	0.000	-.8590574	-.6883209
7	-.0778342	.016932	-4.60	0.000	-.1110204	-.0446479
8	-.8706626	.0271755	-32.04	0.000	-.9239257	-.8173994
totalrooms	.0345704	.0006373	54.24	0.000	.0333212	.0358195
bedrooms	.0327437	.0010952	29.90	0.000	.0305971	.0348903
fullbaths	.228125	.0011964	190.68	0.000	.2257801	.2304698
halfbaths	.1496663	.0011763	127.24	0.000	.1473608	.1519717
fireplaces	.115184	.0011458	100.52	0.000	.1129382	.1174297
bsmtgarage	.0133627	.0008975	14.89	0.000	.0116037	.0151217
_cons	10.47965	.0237628	441.01	0.000	10.43307	10.52622

Source	SS	df	MS			
				Number of obs	=	319,363
				F(54, 319308)	=	11331.19
Model	64225.8417	54	1189.36744	Prob > F	=	0.0000
Residual	33515.8529	319,308	.104964025	R-squared	=	0.6571
				Adj R-squared	=	0.6570
Total	97741.6946	319,362	.306052989	Root MSE	=	.32398

log_fairmark~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
nodedensity	.2761108	.0186542	14.80	0.000	.2395491	.3126724
log_lotarea	.1362518	.0008094	168.34	0.000	.1346655	.1378382
yearbltedited	-.0133199	.0001152	-115.65	0.000	-.0135456	-.0130941
yearbltedited2	.0000664	6.51e-07	101.95	0.000	.0000651	.0000676
style						
2	.0505727	.0034337	14.73	0.000	.0438427	.0573027
3	.0282746	.0029837	9.48	0.000	.0224267	.0341225
4	.0814405	.0019975	40.77	0.000	.0775254	.0853556
5	-.0265903	.0022063	-12.05	0.000	-.0309146	-.0222659
6	.2716633	.0372053	7.30	0.000	.198742	.3445845
7	.1947654	.006989	27.87	0.000	.1810672	.2084637
8	.2594213	.0836703	3.10	0.002	.09543	.4234126
9	.1256485	.0043414	28.94	0.000	.1171395	.1341575
10	-.3632004	.0284761	-12.75	0.000	-.4190127	-.3073881
11	.2477422	.0076315	32.46	0.000	.2327847	.2626996
12	.2772981	.0069945	39.64	0.000	.263589	.2910072
13	-.1347154	.0039726	-33.91	0.000	-.1425015	-.1269292
14	.2913192	.0144404	20.17	0.000	.2630165	.319622
15	.2914812	.0765212	3.81	0.000	.1415018	.4414606

16	.1073596	.0024621	43.61	0.000	.102534	.1121853
17	-.0237379	.035286	-0.67	0.501	-.0928975	.0454216
18	-.0407364	.0048521	-8.40	0.000	-.0502465	-.0312264
19	.2037876	.0104787	19.45	0.000	.1832496	.2243255
20	.128469	.0049644	25.88	0.000	.1187388	.1381991
24	.3689696	.187079	1.97	0.049	.0023002	.7356391
25	.3934557	.1870713	2.10	0.035	.0268013	.7601101
exteriorfinish						
2	.0937112	.0015069	62.19	0.000	.0907577	.0966648
3	.149293	.0050749	29.42	0.000	.1393463	.1592397
4	.1503425	.008129	18.49	0.000	.1344098	.1662752
5	.0341267	.0131433	2.60	0.009	.0083662	.0598871
6	.0868983	.0020104	43.22	0.000	.0829579	.0908387
7	.2707318	.0263163	10.29	0.000	.2191525	.322311
8	-.0154505	.0285933	-0.54	0.589	-.0714925	.0405915
roof						
2	.1989357	.0027636	71.98	0.000	.1935191	.2043522
3	.0804199	.01962	4.10	0.000	.0419654	.1188745
4	.0663583	.0044911	14.78	0.000	.0575558	.0751607
5	.2953521	.0085906	34.38	0.000	.2785147	.3121895
6	.0914146	.0332891	2.75	0.006	.026169	.1566602
basement						
2	.0862564	.013098	6.59	0.000	.0605846	.1119282
3	-.0028854	.0119498	-0.24	0.809	-.0263068	.0205359
4	.0441546	.0064257	6.87	0.000	.0315605	.0567487
5	.0133373	.0057406	2.32	0.020	.0020859	.0245886
condition						
2	-.3162173	.0159993	-19.76	0.000	-.3475754	-.2848592
3	-.5481809	.0159471	-34.38	0.000	-.5794367	-.5169251
4	-.7308991	.016071	-45.48	0.000	-.7623978	-.6994004
5	-.813871	.0175615	-46.34	0.000	-.8482912	-.7794509
6	-.7739807	.0435555	-17.77	0.000	-.8593482	-.6886131
7	-.0777938	.0169322	-4.59	0.000	-.1109804	-.0446072
8	-.8709184	.0271753	-32.05	0.000	-.9241811	-.8176557
totalrooms	.0345816	.0006374	54.25	0.000	.0333322	.0358309
bedrooms	.0326879	.0010954	29.84	0.000	.030541	.0348348
fullbaths	.2281088	.0011965	190.65	0.000	.2257637	.2304539
halfbaths	.1496542	.0011765	127.21	0.000	.1473484	.1519601
fireplaces	.115179	.0011461	100.50	0.000	.1129328	.1174253
bsmtgarage	.0133415	.0008976	14.86	0.000	.0115822	.0151009
_cons	10.6278	.0200207	530.84	0.000	10.58856	10.66704

5.6 Appendix F: Stata Chow F Test Regression Models and Calculations

Less Dense Regression Models:

Source	SS	df	MS			
				Number of obs	=	154,728
				F(54, 154673)	=	3560.64
Model	24785.1044	54	458.983416	Prob > F	=	0.0000
Residual	19938.0572	154,673	.128904574	R-squared	=	0.5542
				Adj R-squared	=	0.5540
Total	44723.1617	154,727	.28904562	Root MSE	=	.35903

log_fairmark~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
linkdensity	.4341099	.0516692	8.40	0.000	.3328394	.5353803
log_lotarea	.1117736	.001536	72.77	0.000	.108763	.1147842
style						
2	.0793659	.0070643	11.23	0.000	.06552	.0932117
3	.0478922	.0067812	7.06	0.000	.0346012	.0611832
4	.0944196	.0035352	26.71	0.000	.0874906	.1013486
5	-.0264908	.003942	-6.72	0.000	-.0342171	-.0187646
6	.2760078	.0607498	4.54	0.000	.1569393	.3950762
7	.2929211	.0135445	21.63	0.000	.2663741	.3194682
8	.411785	.1197292	3.44	0.001	.1771182	.6464517
9	.1767975	.0080685	21.91	0.000	.1609833	.1926116
10	-.3020339	.0736909	-4.10	0.000	-.4464665	-.1576013
11	.1867633	.0090125	20.72	0.000	.169099	.2044276
12	.2044235	.0083982	24.34	0.000	.1879631	.2208839
13	-.1173666	.0056382	-20.82	0.000	-.1284174	-.1063157
14	.2712691	.0198032	13.70	0.000	.2324552	.3100829
15	.3518722	.1469924	2.39	0.017	.06377	.6399743
16	.1034885	.0038845	26.64	0.000	.095875	.1111019
17	.0860592	.0901312	0.95	0.340	-.0905961	.2627146
18	-.0443497	.0080965	-5.48	0.000	-.0602187	-.0284806
19	.2121487	.0166817	12.72	0.000	.1794529	.2448445
20	.1126425	.0063347	17.78	0.000	.1002265	.1250585
24	.4308879	.253919	1.70	0.090	-.0667881	.9285639
25	.4389484	.2073468	2.12	0.034	.0325529	.8453439
exteriorfinish						
2	.1414317	.0024637	57.41	0.000	.1366028	.1462606
3	.2043064	.0076251	26.79	0.000	.1893613	.2192514
4	.1871027	.012903	14.50	0.000	.1618131	.2123922
5	.0817651	.0213929	3.82	0.000	.0398354	.1236948
6	.1391222	.0033803	41.16	0.000	.1324968	.1457476
7	.3517595	.0379087	9.28	0.000	.2774592	.4260597
8	-.0432859	.0848613	-0.51	0.610	-.2096123	.1230404
yearblt	.009847	.0002238	44.00	0.000	.0094084	.0102857
yearbltedited2	.0000554	1.18e-06	46.84	0.000	.0000531	.0000578
roof						
2	.1905695	.0035813	53.21	0.000	.1835502	.1975888
3	.1171609	.0321692	3.64	0.000	.05411	.1802118
4	.0386943	.0053139	7.28	0.000	.0282791	.0491094
5	.3069656	.0109912	27.93	0.000	.2854231	.3285081
6	.0609071	.0509177	1.20	0.232	-.0388905	.1607047
basement						
2	.115963	.0225577	5.14	0.000	.0717503	.1601757
3	.0299799	.0221852	1.35	0.177	-.0135026	.0734623
4	.0516159	.011589	4.45	0.000	.0289017	.0743302
5	.0250839	.0106262	2.36	0.018	.0042568	.045911
condition						
2	-.3008581	.0245854	-12.24	0.000	-.349045	-.2526713
3	-.5827736	.0245025	-23.78	0.000	-.6307981	-.5347492
4	-.7655201	.024652	-31.05	0.000	-.8138376	-.7172026
5	-.8458511	.0264963	-31.92	0.000	-.8977833	-.793919
6	-.9547752	.0680595	-14.03	0.000	-1.08817	-.8213799
7	-.0735065	.0260982	-2.82	0.005	-.1246584	-.0223547
8	-.9104985	.0392614	-23.19	0.000	-.9874501	-.833547
totalrooms	.0270124	.0009592	28.16	0.000	.0251325	.0288924
bedrooms	.0328639	.0015982	20.56	0.000	.0297315	.0359964
fullbaths	.2394016	.0019106	125.30	0.000	.2356569	.2431464
halfbaths	.1672313	.0018731	89.28	0.000	.1635601	.1709026
fireplaces	.1066162	.0018377	58.02	0.000	.1030143	.1102181
bsmtgarage	.0372357	.0016886	22.05	0.000	.0339261	.0405452
_cons	-9.422228	.4436694	-21.24	0.000	-10.29181	-8.552645

Source	SS	df	MS			
				Number of obs	=	154,728
				F(54, 154673)	=	3560.50
Model	24784.6494	54	458.974989	Prob > F	=	0.0000
Residual	19938.5123	154,673	.128907516	R-squared	=	0.5542
				Adj R-squared	=	0.5540
Total	44723.1617	154,727	.28904562	Root MSE	=	.35904

log_fairmark~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
nodedensity	.28232	.0344761	8.19	0.000	.2147475	.3498924
log_lotarea	.1116963	.001536	72.72	0.000	.1086858	.1147069
style						
2	.0793869	.0070644	11.24	0.000	.0655409	.0932329
3	.0478818	.0067813	7.06	0.000	.0345907	.061173
4	.0944277	.0035353	26.71	0.000	.0874986	.1013568
5	-.0264998	.0039421	-6.72	0.000	-.0342262	-.0187734
6	.2760592	.0607505	4.54	0.000	.1569894	.395129
7	.2929895	.0135447	21.63	0.000	.2664421	.3195369
8	.4118342	.1197306	3.44	0.001	.1771648	.6465037
9	.1767884	.0080686	21.91	0.000	.1609741	.1926027
10	-.3018964	.0736917	-4.10	0.000	-.4463307	-.1574621
11	.1868363	.0090127	20.73	0.000	.1691716	.2045009
12	.2044886	.0083984	24.35	0.000	.1880279	.2209493
13	-.1173705	.0056383	-20.82	0.000	-.1284215	-.1063195
14	.2713327	.0198034	13.70	0.000	.2325184	.310147
15	.3519371	.1469941	2.39	0.017	.0638316	.6400426
16	.1034865	.0038845	26.64	0.000	.0958729	.1111001
17	.0861614	.0901323	0.96	0.339	-.090496	.2628188
18	-.0443902	.0080966	-5.48	0.000	-.0602594	-.0285211
19	.2121421	.0166819	12.72	0.000	.1794459	.2448383
20	.1127125	.0063348	17.79	0.000	.1002964	.1251285
24	.4309762	.2539219	1.70	0.090	-.0667055	.9286579
25	.4389914	.2073492	2.12	0.034	.0325913	.8453915
exteriorfinish						
2	.141455	.0024639	57.41	0.000	.1366257	.1462842
3	.2043321	.0076252	26.80	0.000	.1893868	.2192774
4	.1871305	.0129031	14.50	0.000	.1618407	.2124203
5	.0817941	.0213932	3.82	0.000	.0398639	.1237243
6	.1391105	.0033804	41.15	0.000	.132485	.1457361
7	.3517841	.0379091	9.28	0.000	.277483	.4260852
8	-.0433356	.0848622	-0.51	0.610	-.2096638	.1229927
yearblt	.0098448	.0002238	43.99	0.000	.0094062	.0102835
yearbltedited2	.0000554	1.18e-06	46.83	0.000	.0000531	.0000577
roof						
2	.1905743	.0035814	53.21	0.000	.1835549	.1975937
3	.1171844	.0321695	3.64	0.000	.0541328	.180236
4	.03872	.005314	7.29	0.000	.0283048	.0491353
5	.3069855	.0109913	27.93	0.000	.2854427	.3285282
6	.0609478	.0509183	1.20	0.231	-.0388509	.1607466
basement						
2	.1161285	.0225578	5.15	0.000	.0719156	.1603413
3	.0299462	.0221854	1.35	0.177	-.0135368	.0734291
4	.0515623	.0115891	4.45	0.000	.0288478	.0742767
5	.0250259	.0106263	2.36	0.019	.0041987	.0458532
condition						
2	-.3008977	.0245856	-12.24	0.000	-.349085	-.2527103
3	-.5828368	.0245028	-23.79	0.000	-.6308618	-.5348118
4	-.7655646	.0246523	-31.05	0.000	-.8138826	-.7172465
5	-.8459129	.0264966	-31.93	0.000	-.8978456	-.7939802
6	-.9548163	.0680603	-14.03	0.000	-1.088213	-.8214195
7	-.0735201	.0260985	-2.82	0.005	-.1246726	-.0223677
8	-.9105616	.0392618	-23.19	0.000	-.987514	-.8336092
totalrooms	.0270104	.0009592	28.16	0.000	.0251304	.0288904
bedrooms	.0328724	.0015982	20.57	0.000	.02974	.0360049
fullbaths	.2394256	.0019106	125.31	0.000	.2356809	.2431704
halfbaths	.1672391	.0018731	89.28	0.000	.1635678	.1709104
fireplaces	.1066101	.0018377	58.01	0.000	.1030081	.110212
bsmtgarage	.0372173	.0016886	22.04	0.000	.0339076	.0405269
_cons	-9.271286	.442639	-20.95	0.000	-10.13885	-8.403723

More Dense Regression Models:

Source	SS	df	MS
Model	37193.1862	53	701.758231
Residual	12372.1833	164,700	.07511951
Total	49565.3696	164,753	.300846537

Number of obs	=	164,754
F(53, 164700)	=	9341.89
Prob > F	=	0.0000
R-squared	=	0.7504
Adj R-squared	=	0.7503
Root MSE	=	.27408

log_fairmark~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
linkdensity	-5.070106	.0763067	-66.44	0.000	-5.219666	-4.920547
log_lotarea	.151387	.0008748	173.05	0.000	.1496724	.1531017
style						
2	.0463787	.0034648	13.39	0.000	.0395877	.0531697
3	.0065159	.0029393	2.22	0.027	.0007549	.0122768
4	.0760616	.002238	33.99	0.000	.0716753	.080448
5	-.0208223	.0024011	-8.67	0.000	-.0255284	-.0161162
6	.2596067	.0428701	6.06	0.000	.1755823	.3436311
7	.1537497	.0072333	21.26	0.000	.1395725	.1679268
8	.0117496	.1119098	0.10	0.916	-.2075911	.2310904
9	.1100275	.0046663	23.58	0.000	.1008818	.1191733
10	-.4286814	.0266969	-16.06	0.000	-.4810066	-.3763561
11	-.1034235	.0733507	-1.41	0.159	-.2471893	.0403423
12	-.1026096	.0969406	-1.06	0.290	-.292611	.0873919
13	-.2752335	.0066023	-41.69	0.000	-.2881739	-.262293
14	.2703311	.0209691	12.89	0.000	.2292322	.31143
15	.2630986	.0793885	3.31	0.001	.1074988	.4186984
16	.0479691	.0033787	14.20	0.000	.0413468	.0545913
17	-.056841	.0332512	-1.71	0.087	-.1220126	.0083306
18	-.044489	.0055387	-8.03	0.000	-.0553448	-.0336333
19	.1729561	.0124171	13.93	0.000	.1486188	.1972933
20	-.0697866	.0117417	-5.94	0.000	-.0928002	-.0467731
24	.2673753	.2741734	0.98	0.329	-.2699986	.8047493
exteriorfinish						
2	.0666284	.0017623	37.81	0.000	.0631743	.0700825
3	.1187855	.0064152	18.52	0.000	.1062119	.131359
4	.1452248	.0096219	15.09	0.000	.1263661	.1640836
5	-.0001926	.0151747	-0.01	0.990	-.0299347	.0295495
6	.0586014	.0022922	25.57	0.000	.0541087	.0630941
7	.1700232	.0348803	4.87	0.000	.1016585	.2383879
8	-.0478078	.0261716	-1.83	0.068	-.0991035	.003488
yearbltedited	-.0128294	.0001244	-103.13	0.000	-.0130732	-.0125856
yearbltedited2	.0000558	7.41e-07	75.32	0.000	.0000543	.0000573
roof						
2	.1785641	.0047504	37.59	0.000	.1692533	.1878748
3	.0639662	.0225325	2.84	0.005	.0198029	.1081295
4	-.0070798	.0116159	-0.61	0.542	-.0298466	.015687
5	.1766783	.0146556	12.06	0.000	.1479537	.2054029
6	.0927391	.0409338	2.27	0.023	.0125098	.1729684
basement						
2	.0807021	.0145485	5.55	0.000	.0521873	.1092168
3	-.0294285	.0126005	-2.34	0.020	-.0541252	-.0047318
4	.0357325	.0069593	5.13	0.000	.0220924	.0493726
5	-.0018727	.0060806	-0.31	0.758	-.0137905	.0100451
condition						
2	-.3349354	.0195612	-17.12	0.000	-.3732748	-.296596
3	-.5111502	.0194999	-26.21	0.000	-.5493696	-.4729308
4	-.6906123	.0197086	-35.04	0.000	-.7292407	-.6519839
5	-.7786848	.0221733	-35.12	0.000	-.822144	-.7352256
6	-.5895835	.0522937	-11.27	0.000	-.692078	-.4870889
7	-.0837753	.0206331	-4.06	0.000	-.1242157	-.0433348
8	-.7849964	.0363589	-21.59	0.000	-.856259	-.7137339
totalrooms	.0421446	.000801	52.61	0.000	.0405747	.0437146
bedrooms	.0291318	.0014449	20.16	0.000	.0262998	.0319639
fullbaths	.206675	.0014221	145.33	0.000	.2038877	.2094623
halfbaths	.1310356	.001398	93.73	0.000	.1282955	.1337757
fireplaces	.1187861	.0013481	88.11	0.000	.1161438	.1214283
bsmtgarage	.0100098	.0009496	10.54	0.000	.0081486	.011871
_cons	12.51953	.0385068	325.13	0.000	12.44405	12.595

Source	SS	df	MS			
				Number of obs	=	164,635
				F(53, 164581)	=	9382.82
Model	37207.6239	53	702.030639	Prob > F	=	0.0000
Residual	12314.0939	164,581	.074820872	R-squared	=	0.7513
				Adj R-squared	=	0.7513
Total	49521.7178	164,634	.300798849	Root MSE	=	.27353

log_fairmark~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
nodedensity	-3.731193	.052498	-71.07	0.000	-3.834088	-3.628298
log_lotarea	.1502934	.0008743	171.90	0.000	.1485797	.152007
style						
2	.0468153	.0034581	13.54	0.000	.0400374	.0535932
3	.0064439	.0029342	2.20	0.028	.000693	.0121948
4	.0769908	.0022346	34.45	0.000	.0726111	.0813705
5	-.0206225	.0023972	-8.60	0.000	-.0253209	-.0159241
6	.2594883	.0427849	6.06	0.000	.1756308	.3433457
7	.1546406	.0072191	21.42	0.000	.1404913	.16879
8	.0072797	.1116872	0.07	0.948	-.2116249	.2261843
9	.1083896	.0046582	23.27	0.000	.0992595	.1175196
10	-.4302688	.0266439	-16.15	0.000	-.4824903	-.3780473
11	-.1091653	.0732052	-1.49	0.136	-.2526458	.0343153
12	-.1062602	.0967478	-1.10	0.272	-.2958838	.0833634
13	-.2791465	.0065931	-42.34	0.000	-.2920689	-.2662241
14	.2678385	.0209278	12.80	0.000	.2268205	.3088565
15	.2640626	.0792306	3.33	0.001	.1087724	.4193529
16	.0453494	.0033756	13.43	0.000	.0387333	.0519656
17	-.0564255	.033185	-1.70	0.089	-.1214675	.0086165
18	-.0442968	.0055371	-8.00	0.000	-.0551493	-.0334443
19	.1730577	.0123926	13.96	0.000	.1487685	.1973469
20	-.0756606	.0117205	-6.46	0.000	-.0986326	-.0526885
24	.2785533	.2736283	1.02	0.309	-.2577522	.8148589
exteriorfinish						
2	.0685996	.0017612	38.95	0.000	.0651476	.0720515
3	.1211888	.0064062	18.92	0.000	.1086328	.1337448
4	.1445788	.0096029	15.06	0.000	.1257574	.1634002
5	-.0000172	.0151657	-0.00	0.999	-.0297416	.0297072
6	.0602848	.0022894	26.33	0.000	.0557976	.0647721
7	.1672877	.034811	4.81	0.000	.0990588	.2355166
8	-.0480172	.0261197	-1.84	0.066	-.0992111	.0031768
yearbltedited	-.0127923	.0001243	-102.88	0.000	-.013036	-.0125486
yearbltedited2	.000056	7.41e-07	75.58	0.000	.0000546	.0000575
roof						
2	.1794876	.0047428	37.84	0.000	.1701918	.1887835
3	.0629255	.0224878	2.80	0.005	.0188499	.1070012
4	-.0078001	.0115928	-0.67	0.501	-.0305218	.0149217
5	.1765299	.0146265	12.07	0.000	.1478623	.2051975
6	.0930614	.0408525	2.28	0.023	.0129914	.1731313
basement						
2	.084248	.0145491	5.79	0.000	.0557321	.112764
3	-.0301671	.012584	-2.40	0.017	-.0548315	-.0055028
4	.0359196	.0069478	5.17	0.000	.022302	.0495372
5	-.0012864	.0060696	-0.21	0.832	-.0131827	.0106098
condition						
2	-.3334526	.0195224	-17.08	0.000	-.371716	-.2951892
3	-.5095827	.0194612	-26.18	0.000	-.5477263	-.4714391
4	-.6884727	.0196701	-35.00	0.000	-.7270255	-.6499198
5	-.7764514	.0221364	-35.08	0.000	-.8198383	-.7330645
6	-.5892693	.0521897	-11.29	0.000	-.69156	-.4869786
7	-.0820875	.0205929	-3.99	0.000	-.1224492	-.0417258
8	-.7834878	.0362866	-21.59	0.000	-.8546088	-.7123667
totalrooms	.0422727	.0007997	52.86	0.000	.0407052	.0438401
bedrooms	.0290122	.0014426	20.11	0.000	.0261846	.0318397
fullbaths	.206341	.0014197	145.34	0.000	.2035584	.2091236
halfbaths	.1308878	.0013958	93.77	0.000	.1281521	.1336235
fireplaces	.1186404	.001346	88.14	0.000	.1160022	.1212785
bsmtgarage	.0103053	.0009481	10.87	0.000	.0084471	.0121635
_cons	10.86211	.023868	455.09	0.000	10.81532	10.90889

Calculations:

Link Density:

$$F = \frac{(33529.0 - (12372.2 + 19938.1))/58}{(12372.2 + 19938.1)/(154728 + 164754 - 2 \text{ x } 58)} = 207.7$$

Node Density:

$$F = \frac{(33515.9 - (12314.1 + 19938.5))/58}{(12314.1 + 19938.5)/(154728 + 164754 - 2 \times 58)} = 215.7$$