Seattle Housing Prices: Assessing the effects of accessibility to public goods

*Ahmed Almotaileq, Abigail Edelman, Christopher Lacrampe*

**Introduction**

According to S&P/Case-Shiller Home Prices Indices, Seattle has had the fastest growing housing market for over a year with a growth rate of 13.2% - more than double the national average (Rosenberg). When buying or selling a home, a number of factors must be considered to determine the value, including location, square-footage, number of bedrooms and bathrooms, proximity to public goods, etc., and the impact on price of each of these factors is different. This paper will look at the effect that proximity to public goods has on rent prices, and more specifically, the effect of proximity to Link Light Rail stations.

In 2016, an average of 71,058 people rode the Link Light Rail on an average weekday according to the Sound Transit quarterly report for the fourth quarter of 2016, October 1st – December 31st (Sound Transit Quarterly Report). This value is nearly a 10% increase from the previous year, which shows that not only is the light rail a highly valued amenity in the Seattle area, but its popularity as a way to travel in to and out of the city is increasing. All of the currently active light rail stations have been in place since 2009; however, the state of Washington has plans to extend the light rail from Tacoma to Everett and as far east as Redmond (Sound Transit Expansion). This analysis will be beneficial to those who live along the proposed, not-yet-constructed light rail line who may be interested to see whether or not the value of their home will be significantly impacted by the implementation of these new stations.

The data used was collected from Redfin.com on a sample of 350 homes for sale in the Seattle area as of February 20th, 2018, as well as data collected from Seattle.gov’s “My Neighborhood Map” on the light rail, public parks, public schools, and hospitals. The data used in the model includes number of bedrooms, number of bathrooms, property type dummy variables (mutli-family with 5+ residents, single family residential, and townhouse), square feet of the property, square feet of the lot, age of the property, distance to nearest public green space (meters), distance to nearest school (meters), distance to nearest hospital (meters), distance to nearest light rail station (meters), and the distance from the nearest light rail interacted with the second quartile threshold of distance to the nearest station in meters (the determined threshold where distance from the station no longer impacts price). Our regression results show that, at a 1% significance level, each additional meter from the light rail station has a negative impact on the price of a home holding all other variables constant.

**Data**

The data come from Redfin.com, a residential real estate company that provides web-based real estate database and brokerage services, and from Seattle.gov’s “My Neighborhood Map”, which is a collection of data on the public goods and services in a particular area of Seattle. The descriptive statistics for each variable are displayed in the table below. There are 211 observations of 16 different variables.

When originally assessing the descriptive statistics of the data, we noticed some extremely small values for square footage (i.e. a property listing of 1 square foot) and extremely high home prices ($7 million dollars or more). The source of the small property listings was a segment of the house type categorical variables, so we elected to assess only Multi-Family (5+), Residential Family, Co-op, and Townhouse style homes relative to multi-family (2-4) as the omitted values. We also elected to set a price ceiling threshold at $3,000,000, to remove houses with extremely irregular characteristics. Furthermore, we elected to remove observations with null/missing values for key variables. The results after this selection are described in the following paragraphs.

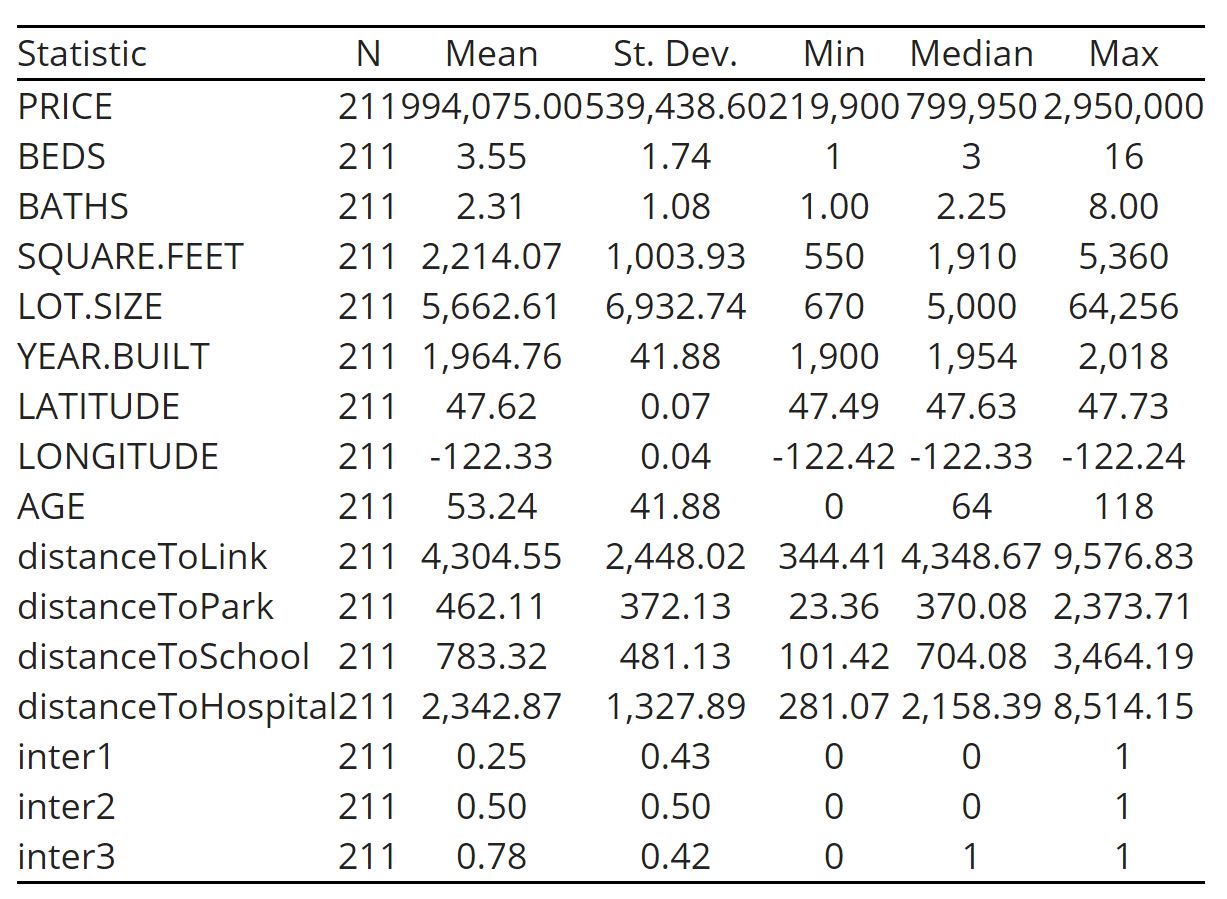
In the first row, the price of homes is shown to have a mean of $994,075 with a median of $799,950 and a standard deviation of $539,438.60. The difference in the mean and the median could indicate that there are some larger prices skewing the mean higher. The lowest price for a home is $219,900, and the highest price is $2,950,000

The second row focuses on the number of bedrooms in the home. The average number of bedrooms is 4 (rounded up from 3.55) with a median of 3 and a standard deviation of 1.7. The minimum number of bedrooms is 1, which makes sense, but the maximum is 16, which means that the maximum may be an outlier. In the third row, the average number of bathrooms in a home is shown to be 2 (rounded down from 2.3) with a median of 2.25 and a standard deviation of 1. The minimum is 1 and the maximum is 8, which makes sense given our 16-bedroom outlier.

Square feet of the property, shown in row four, has an average of 2,214 square feet. The median is 1,910, again indicating that there may be an outlier skewing the mean, and the standard deviation is 1,003 square feet. The minimum number of square feet is 550, and the maximum is 5,360. Square feet of the lot is shown in row five, and that has an average of 5,662.6 square feet with a median of 5,000 and a standard deviation of 6,932.7 square feet. This large standard deviation could be due to the fact that there is a lot of variation in the lot size of different properties, and sure enough, the minimum lot size is 670 square feet and the maximum is 64,256.

We calculated the age variable by taking the difference from the property build-date to the current year. For the age variable in row 9, the average age of the homes was 53.2 years, with a median of 64 and a standard deviation of 41.9 years. The high standard deviation was again due to a wide range in the data, with the minimum age being 0 and the maximum being 118 years old.

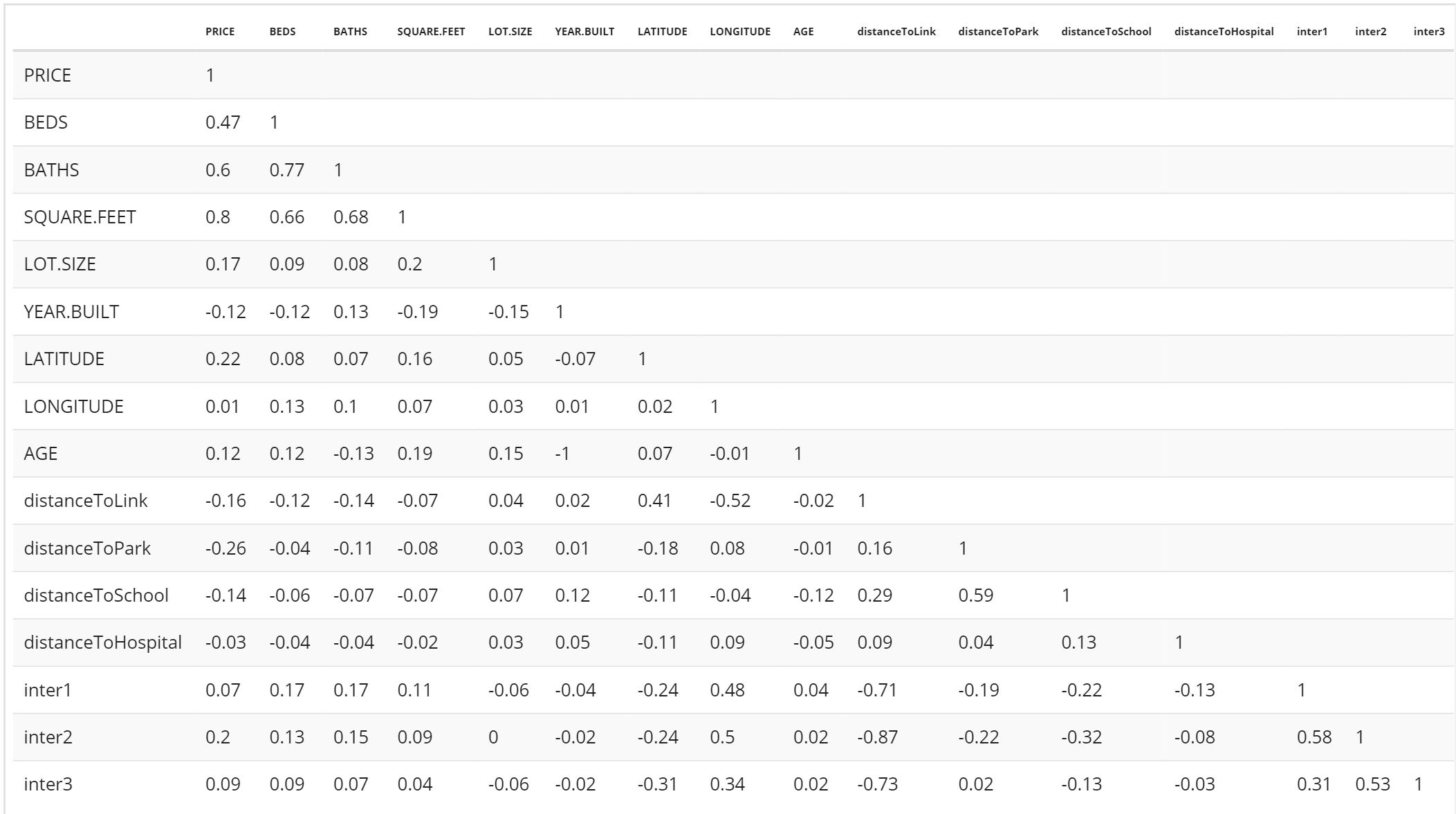
The latitude and longitude (rows 7 and 8) were used to calculate the data in rows 10-16. In order to find the distances between the properties and their closest respective light rail stations, green spaces, schools, and hospitals, the geosphere package in R was used to calculate the distance between the latitude and longitude coordinates of the property and the amenity, keeping in mind the curvature of the earth. The distGeo() function calculates distances by solving Vincenty’s Inverse Problem: where: *s*=*bA*(σ−Δσ), and s (distance) is measured in meters, and accurate to 0.06mm. The function then returns a matrix of distances to each of the public features from the property listing, and by selecting the minimum value of that matrix we identify both the distance to the closest feature, and the name of the feature. Future researchers may elect to calculate distances in a variety of other methods which may include: Euclidian distances by mapping property and public good locations on a 2-demensional plane (Agostini et. al), or by utilizing an alternative geodesic methodology apart from Vincenty’s problem (many options of which the geoDisc package in R provides).



When determining which variables to include in the model, the article titled *The Anticipated Capitalization Effect of a New Metro Line on Housing Prices* written by Claudio Agostini and Gastón Palmucci was referenced. In the article, Agostini and Palmucci are completing a study performed in Chile on the effects of the implementation of a new line of the Santiago subway, and, unlike this paper, they are able to view the short-run effect of the construction and announcements on housing prices as well as the long-run effect of the completed metro line, as opposed to just the long-term effects accessed by this paper. In their analysis, Agostini and Palmucci also use the log of price as their dependent variable. They divide their explanatory variables into 3 groups: first, one that covers structural characteristics like size (in square meters), number of bedrooms and bathrooms, age of the building, etc; the second group covers access to public goods; and the third group covers “dummies equivalent to fixed effects for community, month, and year”, which include a variable for the announcement of their metro line, the release of more information on the basic engineering, and a variable to distinguish between properties within 1,000 meters of the project and properties further than 1,000 meters away.

Assessing correlation:

To ensure that the model avoids the impacts of perfect multi-collinearity we composed a correlation matrix of our numerical variables. Key takeaways from the table include the high correlation between beds and baths, however we elected to include both as Agostini et. al. and Hewett et. al. include the variables in their models, as they are important characteristics in determining property value.



**Empirical Results**

Based on the research by Agostini and Palmucci, the variables included in this model are log of price (dependent variable), number of bedrooms, number of bathrooms, multi-family home 5+ (dummy), single-family residential (dummy), townhouse (dummy), square feet of the property, square feet of the lot, age of the property, distance to the nearest public green space, distance to the nearest public school, distance to the nearest hospital, distance to the nearest light rail station, and an interaction term to include a difference in differences estimator similar to the 1,000 meter threshold included by Agostini et, al. To set our dummy variables, and make them reproducible, we have elected to set them at the 1st, 2nd, and 3rd quartiles of the distance-to-link variable.

For model composition, model 1 represents a base model for estimating the value of a home without the inclusion of proximity to public goods. We then composed model 2 to include public goods with the exception of proximity to rail-station features. Model 3 introduces the proximity to rail-station features, while models 4-6 then include an interaction of the distance thresholds (Set at quartiles 1-3) to include difference of difference estimators produced in research by Agostini et. al. When assessing the differences between each of the models, we have elected to utilize model 5 because, out of models 4-6 that include the interaction terms, this model has a dummy coefficient that is close to zero while also maintaining significance at the 1% level. However, each of the model’s are fairly robust, as many of the coefficients statistical significance and directions hold across each of the composed models, so readers may elect to select a different model.

Here is a description for each variable and an analysis of the results of the model 5:

The dependent variable is**log(price)** in dollars (not thousands), and each variables is either continuous, or relative to an omitted category. So by taking the coefficient of each variables and multiplying by 100, we can calculate the percent change in the price of the property from increasing one numerical value by 1 unit.

**Number of bedrooms** represents the number of bedrooms for that property in integer form. The coefficient for most of the models is significant at most levels, but is fairly small and negative, which seems counter-intuitive and could be indicative of the bathroom variable taking on the effects of the bedroom variable as they are both highly correlated. For model 5, the coefficient is -0.34 and not significant at or below the 10% level.

**Number of bathrooms** represents the number of bedrooms for that property in integer form. The coefficients for this variable are significant at most thresholds, and positive across the models, indicating that as the number of bathrooms increases, the price of the property increases by 13-15% holding all else constant. For model 5 specifically, the constant indicates a 14.6% change per additional bathroom, and is significant at the 1% level.

**Multi-Family home** represents the style of property that includes 5+ units for multiple family living space. The coefficients for this variable are fairly small and insignificant at all thresholds. For model 5, the coefficient is -0.159 and not significant at or below the 10% level.

**Single-Family, Residential** represents the style of property that only includes enough space for one family. The coefficients for this variable are significant at all thresholds and are positive for all of the models, indicating that a home with a single-family style increases the price by about 30-40% holding all other variables constant and relative to the multi-family 2-4 unit style. For model 5, the coefficient indicates that Single-Family, residential style homes are priced 32.9% higher than multi-family (2-4) style homes at the 5% level holding all else constant.

**Townhouse** represents the style of property that includes a factor level for properties listed in the Townhouse category. The coefficients for this variable are significant at all thresholds and are positive for all of the models, indicating that a home with a Townhouse style increases the price by about 40-50% holding all other variables constant and relative to the multi-family 2-4 unit style. For model 5, the coefficient indicates that a Townhouse style home is priced 41.1% higher than a multi-family (2-4) style home at the 5% level, holding all other variables constant.

The **Square Feet of Property** variable captures the effects of increasing the square feet by one on the price. The coefficient is significant at the 1% level and remains fairly consistent at .03% increase in price per an increase of one square foot holding all other variables constant. This statement holds true for the coefficient in model 5.

The **Lot Size (square feet)** variable captures the effects of increasing lot size (different from internal property size) by one square foot. The coefficient for this variable is insignificant, and extremely small for each of the models.

The **Age of Property** variable is an integer style variable that captures the effects of increasing the age of the house on the price. The coefficient is insignificant at all thresholds. For model 5, this coefficient is reported as 0.001 and is not significant at or below the 10% level.

**Distance to nearest public green space** captures the effects of increasing the distance to the nearest public green space by 1 meter. The coefficients are significant at all thresholds and are negative for all of the models that include the variable. Increasing the distance by 1 meter is associated with a .03 or 0.04% decrease in the price of the property holding all other variables constant. For model 5, the coefficient indicates a 0.04% decrease in price of the property per additional meter of distance to the nearest public green space holding all other variables constant.

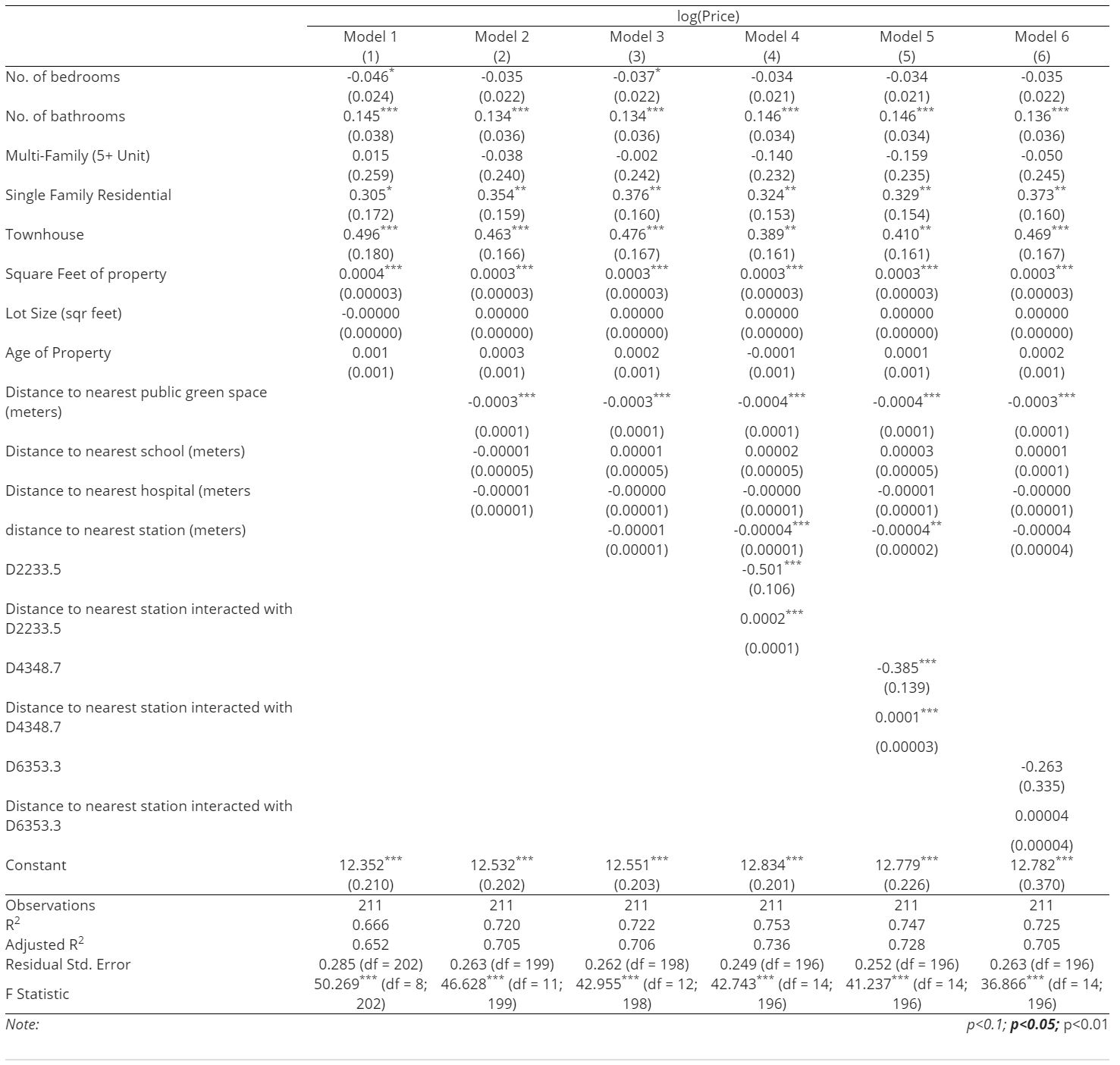
**Distance to nearest public school** captures the effects of increasing the distance to the nearest public school by 1 meter. The coefficients are insignificant at all thresholds and are positive for most of the models that include the variable. Increasing the distance by 1 meter is associated with a .001% increase in the price of the property holding all other variables constant. For model 5, the coefficient is reported as 0.000003 and is not significant at or below the 10% level.

**Distance to nearest public hospital** captures the effects of increasing the distance to the nearest public hospital by 1 meter. The coefficients are insignificant at all thresholds and have negligible effect on housing price. For model 5, the coefficient is reported as -0.00001 and is not significant at or below the 10% level

**Distance to nearest station** captures the effects of increasing the distance to the nearest light rail station by 1 meter. The coefficients are significant at most thresholds and are negative for models 4 and 5. For model 5 increasing the distance by 1 meter is associated with a .0004% decrease in the price of the property holding all other variables constant (this effect is significant at the 5% level).

**D2233.5/4348.7/6353.3** and their **interactions**: Each of these represents an attempt to identify a threshold distance for where the distance to the nearest light rail likely no-longer has an impact on housing price. By selecting a threshold close to 0, we can artificially create a category for property prices that are not affected and are affected by the distance to the light rail. Each of these dummy variables have coefficients that are different from 0 by a large margin, however model 5’s coefficient is the closest to zero while maintaining statistical significance. The coefficient for the interaction is positive, which indicates that a house within 4348.7 meters has a reduced impact of increasing the distance to the nearest station by 1 meter.

The adjusted R-squared for model 5 is 0.747, meaning that, factoring in the number of variables used, the model explains 74.7% of the variation in housing prices. The table below shows the results of the regression models.



**Constant Variance**

To ensure that the inferences made from model 5 are valid and applicable to the population-level (of Seattle), we composed a Breusch-Pagen test of the residuals to confirm our ability to utilize t-tests to assess the statistical significance of the model’s coefficients. The results of the BP test were insignificant, so we can utilize t-tests when assessing the statistical significance of the coefficients, and safely infer the impacts of those coefficients at the population level.

**Conclusion**

In conclusion, this model demonstrates that proximity to a light rail station does have an effect on the pricing of homes in the Seattle market. This means that, as new light rail stations are constructed, home owners close-by can expect an increase in the value of their home as a result of the light rail station opening in the long run.

Other opportunities for research include assessing the impact of accessibility to public transportation on commercial real estate in the Seattle area. Researchers interested in this topic should look to analysis composed by Debrezion et. al. Other researchers might be more interested in accessibility to light rail station in conjunction with the macro-economic factors that shape housing prices—for those inquiries we refer to Li et. al.

Furthermore, future researchers may look to focus on other specific categories of public goods. While this model used distance as a proxy for assessing access and impact of public goods, others may be interested in either conducting an index or using financial valuations to assess the value of the public good, and then maximize that value by the distance (to simulate which public good provides the most value per distance traveled). This may have a more significant impact than only utilizing distance as a proxy. Furthermore, investigating this impact can help shape local policy – perhaps increasing the quality of local parks increases home values, and thus homeowners may be interested in voting for such a bill (although further research would be needed to assess the opportunity costs of voting for such a proposition, and assessing its impact on the overall economy). Another area of expansion upon this study would be to gather a more comprehensive property list, as this dataset contained only 211 observations. Alternatively, light rail stations might have a greater impact on apartment prices, which are not included in the property listings that were collected and utilized for this project.

**References**

Agostini, C. A., & Palmucci, G. A. (2008). The Anticipated Capitalisation Effect of a New Metro

Line on Housing Prices. Fiscal Studies, 29(2), 233-256.

http://web.a.ebscohost.com.proxy.seattleu.edu/ehost/pdfviewer/pdfviewer?vid=1&sid

=07433d5b-2221-4c97-92eb-8e5561d4e7bd%40sessionmgr4009

Debrezion, G., Pels, E. A., & Rietveld, P. (2007). The Impact of Railway Stations on Residential

and Commercial Property Value: A Meta Analysis. SSRN Electronic Journal, 35(2),161-

180. https://link.springer.com/article/10.1007/s11146-007-9032-z

Hewitt, C., & Hewitt, W. (2012). The Effect of Proximity to Urban Rail on Housing Prices in

Ottawa. Journal of Public Transportation, 15(4), 43-65.

http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1080&context=jptn

Li, S., Chen, L., & Zhao, P. (2016). The impact of metro services on housing prices: a case study

from Beijing. Sustainability, 8, 380.

https://smartech.gatech.edu/bitstream/handle/1853/43479/BinYan\_Land+Values+Impacts+of+Subway+Stations.pdf;jsessionid=71FC1FF5925271D11E484D61188E7A3F.smartech?sequence=1

Rosenberg, Mike (2017). Seattle marks a full year as America’s hottest housing market, with no

end in sight. Seattle Times. https://www.seattletimes.com/business/real-estate/seattle-marks-a-full-year-as-americas-hottest-housing-market-with-no-end-in-sight/

Sound Transit Expansion.

https://systemexpansion.soundtransit.org/?\_ga=2.252014400.898339883.1521056418-1580322142.1521056418

Sound Transit Quarterly Report. https://www.soundtransit.org/ridership