Spring 2022

Oregon Counties and Their Effect on Closing Price

Nevaeh Green

Oregon Counties and Their Effect on Closing Price

Nevaeh Green

This study seeks to identify relevant input variables which have an effect of closing prices in dollars in six counties in Oregon. This is done first by collecting sample data from a housing agency regarding houses listing price, closing price, Acres, Square footage, number of bathrooms, number of bedrooms, garage size, and county. Then, using the ordinary least squares method to run regressions and identify the magnitude of the effect of specific regressors on the output variable and using their standard error, an analysis is done both into the statistical and economic significance. This analysis found that when including most variables in the same regression, list price very closely modeled a house's closing price, but none of the other factors were significant. Removing list price, it was found that $acres^2$, $totalSF^2$, and the effect of square footage in several counties were statistically significant, though they were of varying economic significance.

Keywords: Oregon Counties, Housing Prices, Bedrooms, Bathrooms, Square Footage

JEL Codes: C500, D410, E370

Section 1: Introduction

Understanding housing price and its determinants is both incredibly important and incredibly subjective. The difficulty is that people value different things at distinct levels, and quite often when assessing properties people may prefer one property to another for little objective reason. At the same time, there are numerous values of properties which can be objectively measured, things such as its size in square feet, the number of bedrooms, the size of the property, etc. It is these determinants correlated effect on housing price which this study seeks to quantify. According to a U.S. Census Bureau report from 2013, the proportion of families who could afford "moderately priced houses" had decreased by approximately 10% in the last quarter century (Wilson & Callis, 2013). This is a potentially startling trend and quantifying the determinants of a property's value would allow for a greater understanding of the potential cause of this trend if it were due to house's price rather than other outside factors.

In the most general terms, this study seeks to understand the correlation between specific input variables and the output variable of housing prices. The research question could be understood explicitly as, "How does a property's county, acreage, square footage, number of bedrooms, number of bathrooms, garage size, and year constructed affect a house's closing price?" This study is constrained to the state of Oregon, and specifically six counties, so it could be understood as seeking to understand most specifically the effect of a house's location based on county on its closing price.

It may be difficult to conceptualize the importance of this specific question regarding counties in Oregon and even generally when considering all the listed input variables. Since it is not as though you can transpose a property from one county to another to increase its value, why should the effect of its location on its price matter? It is of value for two primary reasons. The first of these is that by understanding the value added or taken away from a property depending on its county, we may open the possibility to create a stronger understanding of the characteristics of each of the counties of interest. Also, though there are of course legitimate concerns regarding using regression results for predictive circumstances, having a quantified value of the effect of these determinants on housing price, we may be able to assess what properties may be of most value in the future.

This is of course not the first study of its kind; numerous studies and analysts have sought to understand housing markets and their determinants. One of these studies seeks to analyze the effect of monetary policy changes on regional housing prices in the United States, and this study found that different regions responded in different degrees to monetary policy shifts and that these responses are related to local regulatory environments (Fischer, Huber, Pfarrhofer, & Staufer-Steinnocher, 2019). However, though there are studies which take up similar goals, this study is much more focused to the state of Oregon specifically and is concerned not with policy measures' effects on housing prices but on variables within the property itself.

Interestingly, this study found that there is a widely differing effect that each of the counties of interest have on the pricing of properties. Several of the regressions found that there were counties which have a statistically significant negative correlation with housing price at a 1% significance level but found that no counties had a statistically significant positive correlation with housing price at any conventional significance levels. Meanwhile, other regressors varied in the magnitude and significance of their correlation with housing price.

Section 2: Data Overview

This data consists of eight input variables and one output variable which will be defined here, and their summary statistics are in *Table 1*:

ListPrice – continuous variable measured in real 2022 dollars

ClosePrice – continuous variable measured in real 2022 dollars

County – discrete variable where for each data point it is one of six counties, Clackamas, Clark, Washington, Multnomah, Yamhill, or Columbia

Lot Size – continuous variable measured in acres of the property which the house is part of

TotalSF – the living space in a house measured in ft^2

Bedrooms – discrete variable, which is either one, two, three, or four bedrooms

YearBuilt – discrete variable for the year of the house's construction

Garage – discrete variable, which is either zero, one, two, three, or four depending on the number of cars which it can fit, zero being that there is no enclosed garage

TotalBaths – discrete variable, which is either one, two, or three, where one is one bathroom, two is one and a half bathrooms, and three is two full bathrooms

Variables	**Mean**	**Median**	**Standard Deviation**	**Minimum**	**Maximum**
List Price	$409,600	$405,500	$59,325	$249,900	$650,000
Close Price	$422,800	$420,000	$63,716	$256,000	$650,000
Acres	0.1770 acres	0.170 acres	0.0610 acres	0.06 acres	0.4800 acres

Total SF	1208 $ft.^2$	1212 $ft.^2$	116.4 $ft.^2$	1000 $ft.^2$	1400.00 $ft.^2$
Bedrooms	2.870	3.000	0.4250	1.000	4.000
Year Built	1974	1975	22.24	1900	2021
Garage	1.623	2.000	0.6610	0.000	3.000
Total Baths	2.350	3.000	0.8740	1.000	3.000

Table 1: Summary statistics for the regressors and regressand

No perfect multi-collinearity is ensured by using base groups for any Boolean variables and their corresponding interaction terms. Imperfect multi-collinearity should also be avoided since the selected base groups were values which made up a sizable proportion of the data. Interestingly, there is evidence that this data is homoscedastic rather than heteroscedastic. When graphing the residuals as a function of predicted house price, the residuals are randomly distributed meaning that there is not greater variation at more extreme values. So, this study will use homoscedasticity-only standard errors to assess the statistical significance of estimated coefficients. Theoretically this makes sense since houses which are more expensive have no reason to have a greater variation in their determinants than houses which are less expensive. There may be exceptional circumstances which causes more expensive houses to have over-inflated values, but this is not necessarily more common in more expensive houses as smaller or older houses may also have historic significance for example which would cause greater variance in their price.

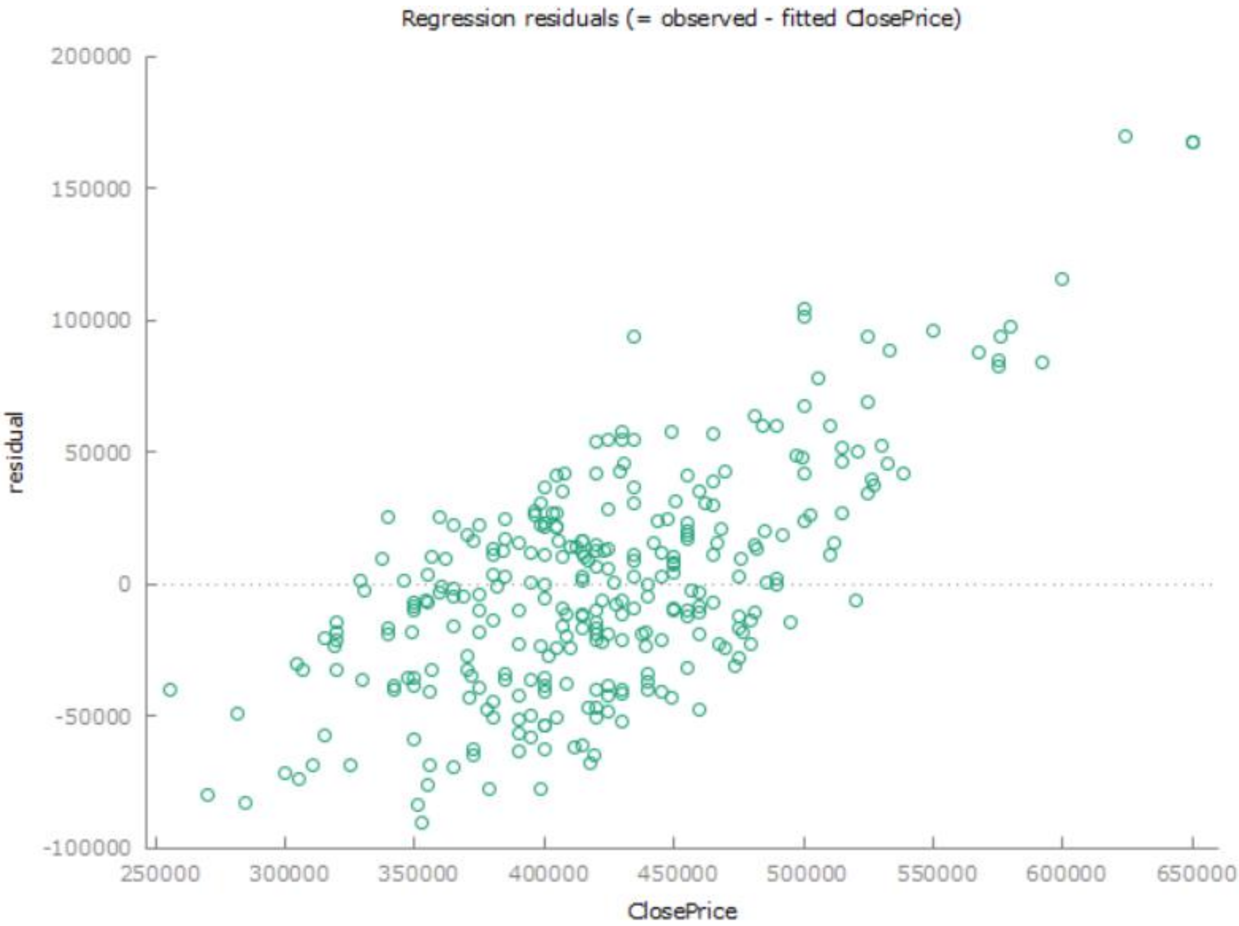

Figure 1: Scatterplot of residuals as a function of ClosePrice, justifying use of homoscedasticity-only standard errors.

However, there are two assumptions which it is much more questionable if they can be assumed to be true for this data. Firstly, there is independent error terms. This is in question since there is correlation between several input variables as generally, the larger a house or the greater its square footage, the more rooms it will have, which is specifically applicable to the regressors bathrooms and bedrooms. This assumption and the assumption of the normality of error terms are immediately drawn into question because of the origins of the data used in this study. This data was sourced from an acquaintance of a contributor for the study rather than from a published source. This data is of course not randomly generated and is taken from properties which they have interacted with as a seller of real estate. But this is not a randomly drawn sample of houses in Oregon since most counties are not represented in this sample and could only be understood as a randomly drawn sample of houses which this specific real estate company represents.

This shortcoming may be addressed by limiting the scope and generalizability of this study. However, another more fundamental problem is that there seems to be an equivalent number of houses from each of the six counties which would be unlikely in a properly randomly drawn sample since each county certainly varies in the gross number of homes it has. The range of several values is also limited, particularly for square footage of the houses, as there is a range of only approximately five hundred square feet which means either the sample is too small and is not representative of the population of properties in these counties, or it has not been randomly drawn from all properties. Because of these shortcomings, there is little to no external validity given the specific circumstances of these six Oregon counties and houses in these counties represented by this real estate group, and their internal validity is questionable again because of the failure to satisfy that the data is randomly and independently sampled.

Section 3: Methodology

To determine the best form for the data for each regressor, they were analyzed visually in a scatterplot with the regressors serving as the independent variable and housing price as the dependent variable. This allowed for visual analysis of whether there is a linear relationship between these values, and for most regressors there was determined to be a linear relationship. However, the relationship between acres and house price and square footage and house price was not able to be definitively determined using this method. To further support decisions regarding the form of the data are going to be used, a simple linear regression was created with each of the continuous input variables and their square and log forms, and the close price was the regressor, and the assessment would be made based on their adjusted R^2 to see which has the most explanatory value. For most of these variables there was either no difference or the non-linear forms of the variables had less explanatory value. But, for acres and total square feet, each of the quadratic forms of the variables had higher explanatory value, so based on this and the x-y scatterplot results the quadratic form was used for these variables.

This study includes quite a few discrete variables so, to properly regress them and analyze the effect that each of them has, a number of Boolean variables were created for each discrete variable depending on the range of values. A binary variable was created for each of the six counties, one for each of the possible values of number of bedrooms, for the size of the garage, and for the number of bathrooms. With each of these dummy variables, there is a question of whether they have differing slopes as well as intercepts that model their correlation with the continuous variables and the output variable. To account for this, interaction terms were created between each of the Boolean variables and each discrete variable. These are of varying statistical and practical significance and will be discussed more in depth in the results section analyzing the outputs from the ordinary least squares method of multiple regression.

Theoretically it would be expected that all these variables except for purchase year would have a positive correlation with the price of the house, the only variables which may not are the binary variables for county as specific counties may be correlated with a decrease in a house's value while others may be correlated with an increase in a house's value. By running multiple regressions with different numbers of regressors, the effect which each input variable has will not only be found to be positive or negative but will also be quantified. As discussed, there are binary variables and interaction terms which will be used in different regressions, but the simplified population formula can be understood as follows:

$$ClosePrice = \beta_0 + \beta_1 ListPrice + \beta_2 Acres + \beta_3 TotalSF + \beta_4 Bedrooms + \beta_5 YearBuilt + \beta_6 Garage + \beta_7 TotalBaths + \beta_8 County$$

Section 4: Results

Regressors	*Regression 1*	*Regression 2*	*Regression 3*	*Regression 4*
Constant	233984	702981**	316351***	386958
	(146925)	(302511)	(29064.1)	(402034)
List Price	0.9268***			
	(0.0304)			
Clark Co.	-5289.61	-39490.6***	-38921.5***	
	(4592.86)	(9221.1)	(9107.15)	
Washington Co.	1877.28	17199.8*	19107.5**	
	(4502.69)	(9263.79)	(9032.7)	
Multnomah Co.	-10826.7**	-17757.8*	-17004.7*	
	(4521.87)	(9349.84)	(9178.24)	
Yamhill Co.	-17172.4***	-71697.7***	-70804.0***	
	(4780.86)	(9180.4)	(9087.53)	
Columbia Co.	-13351.7***	-87255.5***	-86670.3***	
	(5015.23)	(9091.88)	(8985.56)	
Year Built	-83.8493	-178.117		
	(73.7706)	(152.594)		
Lot Size	-14794.7	325365***	358307***	
	(50421.1)	(101803)	(97717.6)	

Total SF				420.366***
Total SF²	0.0021	0.0464***	0.04336***	-0.1054***
	(0.0051)	(0.01007)	(0.0096)	
1.5 Bathrooms	4299.79	20906.9**	22895.8**	
	(4548.23)	(9348.56)	(9326.63)	
2 Bathrooms	4654.98	15499.7**	16082.3**	
	(3150.66)	(6481.19)	(6391.14)	
No Garage	-40518.9*	-45207.4		
	(22074.3)	(45699.5)		
1 Car Garage	-33382.3	-24860.4		
	(21753.6)	(45032.9)		
2 Car Garage	-32218.8	-22272.2		
	(21533.1)	(44575.0)		
2 Bedrooms	11125.5	28340.5	38436.3	54212.0
	(13289.8)	(27489.2)	(26654.3)	(55018.8)
3 Bedrooms	9217.64	40245.1	52832.3**	74257.6**
	(13492.2)	(27853.5)	(26057.1)	(35160.9)
4 Bedrooms	9409.28	21481.7	31770.3	48997.0
	(15704.9)	(32503.7)	(31300.1)	(41765.2)
R^2	0.895	0.549	0.541	0.120

Table 2: Results for regressions 1-4. For each regression, n=300.

Regressors	*Regression 5*	*Regression 6*	*Regression 7*	*Regression 8*
Constant	8766360	2794980***	1974450***	

	(123141100)	(658647)	(600520)	
Clark Co.	-2482500*	-2456190**	-2403930**	-89049.3*
	(1319780)	(1076950)	(1033320)	(53291.4)
Washington Co.	-3844330***	-4191310***	-3486260***	-37500.6
	(1301590)	(1149790)	(1143150)	(-37500.6)
Multnomah Co.	-887077	-727573	-1195450	32689.4
	(908757)	(812349)	(772027)	(47898.9)
Yamhill Co.	-1762550**	-1937720**	-1422770*	-55528.6
	(862906)	(773711)	(776140)	(46577.9)
Columbia Co.	-2036000**	-2519010***	-2215080***	-137231.0***
	(927802)	(765129)	(752168)	(47189.4)
Year Built	648.326	-1202.18***	-788.303***	
	(1691.70)	(326.931)	(301.485)	
Acres				-104766
				(166474)
Acres^2	-2.02247*10^8	346451**	389519***	
	(2.5550*10^8)	(147006)	(146268)	
Total SF				608.013***
				(79.121)
Total SF^2	-4.90573	0.0378195**	0.04694***	-0.2167***
	(6.31952)	(0.018156)	(0.010014)	(0.05092)
1.5 Bathrooms	963178	-48959.5		24354.0**
	(988941)	(46515.1)		(9559.78)

2 Bathrooms	113896	11959.8		14596.6**
	(638863)	(31137.9)		(6515.78)
No Garage	-3324820**	-2952640***		-15207.78
	(1330840)	(991311)		(10885.4)
1 Car Garage	-199370	-375132		-150.517
	(776424)	(706899)		(8235.9)
2 Car Garage	-40030.4	-45023.1		20747.52
	(44013.9)	(43166.1)		(7251.5)
2 Bedrooms	-5632980	-2042.38		20697.3
	(12506900)	(29303.5)		(27866.5)
3 Bedrooms	-6437990	3962.08		29518.9
	(12371000)	(29724.6)		(27911.8)
4 Bedrooms	-9631340	-18129.6		9677.18
	(12144100)	(33851.5)		(32466.9)
*$Acres^2$*Clark*	-262662	-176467	-102496	
	(216320)	(180466)	(177002)	
*$Acres^2$*Washington*	-56147.5	95908.4	116646	
	(157973)	(134648)	(133582)	
*$Acres^2$*Multnomah*	-422091***	-388491***	-310045**	
	(150566)	(133937)	(127313)	
*$Acres^2$*Yamhill*	-308153*	-211924	-236867	
	(171839)	(149440)	(151017)	
*$Acres^2$*Columbia*	36060.2	109332	156966	

	(136190)	(117733)	(117710)	
*TotalSF²*Clark*	0.004004			0.03261
	(0.03690)			(0.03407)
*TotalSF²*Washington*	0.004288			0.03695
	(0.035216)			(0.03139)
*TotalSF²*Multnomah*	-0.03406			-0.03566
	(0.03501)			(0.03212)
*TotalSF²*Yamhill*	-0.02275			-0.01157
	(0.03524)			(0.03121)
*TotalSF²*Columba*	0.02961			0.03440
	(0.03681)			(0.03173)
*YearBuilt*Clark*	1257.58*	1239.86**	1208.59**	
	(660.760)	(535.649)	(514.398)	
*YearBuilt*Washington*	1957.22***	2122.04***	1765.76***	
	(663.647)	(578.767)	(575.451)	
*YearBuilt*Multnomah*	496.284	386.901	621.420	
	(464.438)	(412.723)	(393.122)	
*YearBuilt*Yamhill*	900.168**	963.536**	704.263*	
	(441.994)	(389.347)	(390.516)	
*YearBuilt*Columbia*	962.479**	1220.74***	1061.76***	
	(476.548)	(384.352)	(377.952)	
*Acres²*Beds2*	203070000			
	(255459000)			

*Acres^2*Beds3*	202910000		
	(255492000)		
*Acres^2*Beds4*	202308000		
	(255537000)		
*TotalSF^2*Beds2*	4.96539		
	(6.32275)		
*TotalSF^2*Beds3*	4.93765		
	(6.32279)		
*TotalSF^2*Beds4*	4.93675		
	(6.32268)		
*YearBuilt*Beds2*	-2052.37		
	(1760.13)		
*YearBuilt*Beds3*	-1617.19		
	(1697.68)		
*Acres^2*Garage0*	-273046		
	(540810)		
*Acres^2*Garage1*	-418665		
	(353353)		
*YearBuilt*Garage0*	1659.88**	1471.30***	-40.5311*
	(678.943)	(505.483)	(22.7173)
*YearBuilt*Garage1*	78.8887	163.542	-24.6381
	(396.506)	(358.670)	(22.3811)
*YearBuilt*Garage2*			-21.4995

			(22.0954)	
*Acres²*Baths2*	-363363			
	(338158)			
*Acres²*Baths3*	74545.3			
	(318984)			
*TotalSF²*Baths2*	0.05944*	0.04600		
	(0.03401)	(0.03068)		
*TotalSF²*Baths3*	0.00829	0.003964		
	(0.02599)	(0.02161)		
*YearBuilt*Baths2*	-515.961			
	(501.964)			
*YearBuilt*Baths3*	-55.5873			
	(325.983)			
R^2	0.625	0.607	0.574	0.558

Table 3: Results for regressions 4-6. For each regression n=300.

The first three regressions whose results are displayed in *Table 2* focuses on the previously defined variables in the data overview section of this study. The first regression has the highest R^2 value out of the six regressions, meaning that it has the best explanatory value for the change in housing price based on the change in independent variables. However, only a minority of the regressors are statistically significant at any conventional significance levels. For these reasons in the following regressions the variable *ListPrice* was excluded. This is because *ListPrice* is inherently highly correlated with the dependent variable *ClosePrice*. Using Gretl to create a correlation matrix, the correlation between *ListPrice* and *ClosePrice* is equal to 0.94. This value is remarkably close to one which would mean that these two variables were perfectly correlated. To ensure that there is a sufficient difference of means between these two variables, a hypothesis test is constructed in *Equation 1*. This test finds statistically significant evidence that there is a difference of the means at all conventional significance levels, so this relationship is still worth understanding since it is a determinant of the regressand; however, since it is so highly correlated with the output variable, it is excluded from the rest of the regressions since it being included nullifies the statistical significance of most other variables.

$$H_0: \mu_{ListPrice} - \mu_{ClosePrice} = 0$$

$$H_1: \mu_2 - \mu_1 \neq 0$$

$$SE_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{59325^2}{300} + \frac{63716^2}{300}}$$

$$= \sqrt{11731518.75 + 13532428.85} = 5026.33$$

$$\mu_2 = 422800 \; \mu_1 = 409600 \; \mu_2 - \mu_1 = 13200$$

$$5026.33 * 1.645 = 8268.31 \; 13200 \pm 8268.31 = (4931.69, 21468.31)$$

$$5026.33 * 1.96 = 9851.60 \; 13200 \pm 9851.60 = (3348.40, 23051.60)$$

$$5026.33 * 2.575 = 12942.79 \; 13200 \pm 12942.79 = (257.211, 26142.79)$$

Equation 1: Difference of means hypothesis test for ListPrice and ClosePrice. There is statistically significant evidence that $\mu_2 \neq \mu_1$ at all conventional significance levels, so we reject the null hypothesis and support the alternative hypothesis that there is a difference of the list price and closing price means.

In regression one there are three regressors with statistically significant coefficients at each of the standard significance levels, one variable which is statistically significant at the 5% and 10% levels, and finally one more input variable which is only statistically significant at the 10% significance level. The variable *ListPrice* has already been discussed at length for its value in this study, but a final note is that it is of economic significance since an increase in *ListPrice* of $1,000 is correlated with an increase in *ClosePrice* of $926.82. Interestingly, three of the counties have a statistically significant coefficient at the 5% level, while not having a garage is only significant at the 10% significance level. Each of these variables which are statistically significant are also of economic significance since a house being in Multnomah County is correlated with a decrease in its *ClosePrice* of $10,826.70, being in Yamhill County is correlated with a decrease of $17,172.40, being in Columbia County is correlated with a decrease of $13,351.70, and having no garage is correlated with a decrease in value of $40,518.90.

Interestingly, three variables which would traditionally be presumed to have a significant effect on house price (square footage, number of bathrooms, and number of bedrooms) were found to not have a statistically significant coefficient in this regression. Only having three bathrooms approaches being statistically significant at a 10% significance level. This may be explained by the inclusion of the variable *ListPrice* since intuitively a house will already have a higher list price if it has more square feet, more bedrooms, and more bathrooms. Since this effect is captured by *ListPrice*, there is not enough of an effect from these variables which is not already included in *ListPrice*, and thusly in the first regression none of them have a statistically significant effect on the output variable. The model for the relationship between these input variables and the output variable is as follows:

$$\begin{aligned}\widehat{ClosePrice} = {} & 233984 + 0.927ListPrice - 5289.61Clark + 1888.28Washington \\ & - 10862.7Multnomah - 17172.4Yamhill - 13351.7Columbia \\ & - 83.849YearBuilt - 14794.7Acres^2 + 0.0021TotalSF^2 \\ & + 4299.79Baths2 + 4654.98Baths3 - 40518.9Garage0 \\ & - 33382.3Garage1 - 32218.8Garage2 + 11125.5Beds2 \\ & + 9217.64Beds3 + 9409.28Beds4\end{aligned}$$

Regression two included the same variables as regression one; however, it removed the variable *ListPrice* to increase the statistical significance of the other input variables. This time, all the counties have a statistically significant coefficient at the 10% significance level and three of them are statistically significant at all conventional levels of significance. Once again each of these are also practically significant, and interestingly, now one of the counties (Washington County) is correlated with an increase in *ClosePrice*. What is key to interpreting these coefficients relating to county is that Clackamas County is the base group, so a house in Clackamas County would have a value $702,981 *ceteris paribus*, while a house in Washington county would have a value of $702,981+$17,199.80 or $720,180.80, and houses in other counties would have the value of one in Clackamas county minus their associated coefficient. In this regression, many of the traditionally assumed to be significant determinants have statistically significant coefficients such as square footage, acreage, and number of bathrooms.

> The variable acreage is of questionable practical significance since under most circumstances homeowners are not purchasing land adjacent to their property and adding to its size. However, when normalizing the coefficient to understand a change in price associated with an increase in 0.01 acres, it becomes obvious that this is of practical significance since this is correlated with an increase in price of $3,253.65. Square footage is interesting because it easily passes all significance tests but is not of practical significance since an increase of 100 ft^2 is correlated with an increase in price of $4.64, which typically houses are not priced that specifically down to the dollar and cents. The number of bathrooms is statistically significant at a 5% significance level and is also of practical significance since having 1.5 bathrooms instead of one is correlated with an increase in price of $20,000 and having two bathrooms instead of one is correlated with

an increased price of $15,000. Based on regression two, the model for *ClosePrice* is as follows:

$$\widehat{ClosePrice} = 702981 - 39490.6Clark + 17199.88Washington - 17757.8Multnomah - 71697.7Yamhill - 87255.5Columbia - 178.117YearBuilt + 325365Acres^2 + 0.0464TotalSF^2 + 20906.9Baths2 + 15499.7Baths3 - 45207.4Garage0 - 24860.4Garage1 - 22272.2Garage2 + 28340.5Beds2 + 40245.1Beds3 + 21481.7Beds4$$

Regression number three is the most refined model which does not include interaction terms. For this regression, all but two of the regressors are significant at the 10% significance level, and all but one of those which are significant are also significant at the 5% significance level. Comparing the results from regressions two and three, the coefficients which are produced for common variables are of course different, but it is by a small magnitude. The main difference is that by eliminating the Boolean variables for garage size the statistical significance of the remaining variables was increased to the point where nearly all the regressed input variables are statistically significant. Also, when comparing the R^2 values for the first three regressions, each of the generated models approximate the relationship between the regressors and regressand accurately. Of course, the first regression has the highest explanatory value because of the high degree of correlation between *ListPrice* and *ClosePrice*, between regressions two and three there is a minimal decrease in the explanatory value while once again increasing the statistical significance of almost all the coefficients. Though the model is similar, it is still worth nothing the new model for regression three:

$$\widehat{ClosePrice} = 316351 - 38921.5Clark + 19107.5Washington - 17004.7Multnomah - 70804.0Yamhill - 86670.3Columbia + 358307Acres^2 + 0.04336TotalSF^2 + 22895.8Baths2 + 16082.3Baths3 + 38436.3Beds2 + 52832.3Beds3 + 31770.3Beds4$$

Regressions five through eight vary in a key way from the previous three regressions, these regressions introduce interaction terms between the binary and discrete variables to analyze if there is a difference in the slope of the continuous variables. Regression five included over fifty input variables, of which only a minority are statistically significant. For interaction terms which were nowhere near significant, these were removed in regression six, and after regression six more of the non-significant variables were once again removed until this model was produced from regression seven:

$$\widehat{ClosePrice} = 1974450 - 2403930Clark - 3486260Washington - 1195450Multnomah - 1422770Yamhill - 2215080Columbia - 788.303YearBuilt + 389519Acres^2 + 0.0469TotalSF^2 - 102496Acres^2 * Clark + 116646Acres^2 * Washington - 310045Acres^2 * Multnomah - 236867Acres^2 * Yamhill + 156966Acres^2 * Columbia + 1208.59YearBuilt * Clark + 1765.76YearBuilt * Washington + 621.42YearBuilt * Multnomh + 704.263YearBuilt * Yamhill + 1061.76YearBuilt * Columbia - 40.5311YearBuilt * Garage0 - 24.6381YearBuilt * Garage1 - 21.4995YearBuilt * Garage2$$

Not all these coefficients are statistically significant, the interaction term between *YearBuilt* and garage size is only significant at the 10% significance level for a garage of size zero. The interaction term between $Acres^2$ and *County* are also only statistically significant at a

5% significance level in the case of Multnomah County. The results from regressions four through six imply that there is not a different slope for different values of the discrete variables. For some individual interaction terms there is a statistically significant different slope, but this is likely due either to variation or a sampling error since this was not consistent with other interaction terms. Based on these results, regression three would be the best model for house price other than regression one which includes list price. Based on these results, the counties Clark, Multnomah, Yamhill, and Columbia are correlated with a decrease in housing value from the value of properties in Clackamas County, while Washington county is correlated with an increase in house price from Clackamas County.

Section 5: Conclusion

This study sought to analyze the relationship between Oregon county and house price as well as other potential determinants of house price. By using the ordinary least squares method of multiple regression and the software package Gretl, models were formed based on the data set n=300 which was sourced from an agent working in the housing industry. This study also uses homoscedastic errors based on the justification of the scatter plot of residuals as a function of the output variable.

Regressions one and four included the most input variables, then, using the results, variables which were not of statistical significance were eliminated to best quantify and model the relationship between determinants and house price. Interestingly, this method found that the year the house was built, and the size of the garage had little explanatory value for the output variable based on the results from regressions one through three. Then, beginning with regression five this time using a great deal of interaction terms between the binary and discrete variables, the findings from the final regression seven found that there is limited evidence to suggest that the slope modeling the relationship between regressors and the regressand varies depending on the value of the different discrete variables. Unfortunately, there are fundamental questions about external validity for this study, to the point where it can almost certainly be said to only have internal validity, which itself is questionable due to the sourcing of the data.

To complete the purpose of this study to understand and quantify the effects of different Oregon counties on house price, they will be ranked in terms of the size of their correlation in regression eight since this was the most refined equation used. Firstly, the county which is most negatively correlated with housing price is Columbia County with a correlated negative effect of just over $137,000. Clark county had the second largest correlated effect, with an also negative correlated effect of approximately $89,000. Yamhill county is next with a correlated negative effect of $55,5000 followed by Washington county with a negative effect of $37,500. Clackamas county was the base group, so all these negative correlations are in comparison with the price of homes in Clackamas County, and only Multnomah County had a correlated positive effect over Clackamas County with a correlated positive effect of just over $32,500.

References

Fischer, M. M., Huber, F., Pfarrhofer, & M., Staufer-Steinnocher, P. (2019). The dynamic impact of monetary policy on regional housing prices in the United States. *Real Estate Economics*, *49*(4), 1039-1068.

Wilson, E., & Callis R. R. (2013). *Who could afford to buy a home in 2009? Affordability of buying a home in the United States*. (Current Housing Reports No. H121/13-02) : US Department of Commerce, Economics and Statistics Administration