Regression Analysis For Real Estate Properties

Serena Parve

EMSE 6574

December 12, 2022

A number of factors can be used to determine the price of a real estate property. In this report a dataset of 80 properties with 10 features is used to perform regression analysis to describe the relationship between the price and the features of the property. Table 1 depicts a sample subset of 10 of these 80 properties of the dataset and their features.

Table 1: Sample Subset of the dataset

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Property	PRICE	bedrooms	bathrooms	sqft_living	sqft_lot	floors	Numbers of times viewed	Quality Grade	sqft_above	sqft_basement	Built or Renovated
1	$ 440,000.00	3	2.5	1910	66211	2	0	7	1910	0	1997
2	$ 213,000.00	2	1	1000	10200	1	0	6	1000	0	1961
3	$ 563,500.00	4	1.75	2085	174240	1	0	7	1610	475	1964
4	$ 1,550,000.00	5	4.25	6070	171626	2	0	12	6070	0	1999
5	$ 1,600,000.00	6	5	6050	230652	2	3	11	6050	0	2001
6	$ 350,000.00	3	2.25	1580	47916	1	0	7	1580	0	1979
7	$ 540,000.00	3	2.25	2000	217800	2	0	8	2000	0	1996
8	$ 535,000.00	3	1	1330	40259	1	0	7	1330	0	1977
9	$ 600,000.00	2	2.5	2410	102366	1	0	7	1940	470	1989
10	$ 275,000.00	3	1	1370	17859	1	0	7	1150	220	1930

We perform regression analysis to create a simple model that can explain the relationship between the dependent variable (in this case the price) and the independent variables. The aim is to create a parsimonious model with the least amount of explanatory variables while also adequately describing the variation in the dependent variable.

Initially, we plot a histogram of the Price (which is the dependent variable in this case study). We’d like to see symmetry in the histogram, since having symmetry in the dependent variable would also give us symmetry in the residuals. And we’d like to have residuals be normally distributed – which is also symmetric. We would like the residuals to be normally distributed so we can run statistical tests on them.

From Figure 1, which depicts the histogram of Price the following observations can be made:

- The graph is not symmetric
- The second bar on the left is particularly high skewing the graph
- The histogram is skewed to the left
- Fitting a normal distribution on this could possibly give us

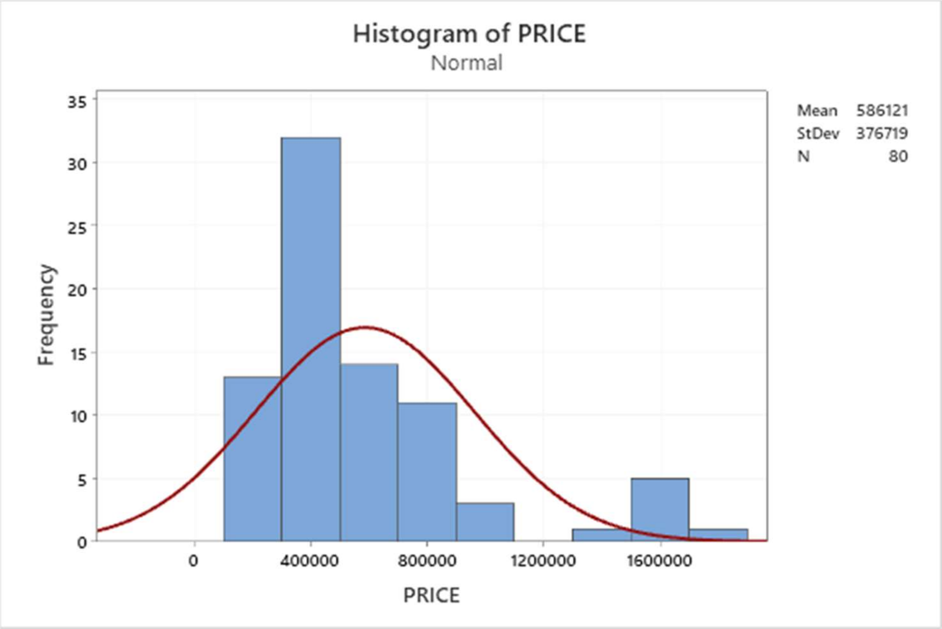


Figure 1: Histogram of Price (Y)

We can further conclude that Price is not normally distributed from the following observations made from Figure 2, which depicts the Probability Plot of Price:

- The points do not lie on a straight line
- Many points lie outside the confidence boundaries of the probability plot
- The p value of the probability plot is very low
- The standard deviation of Price is very large

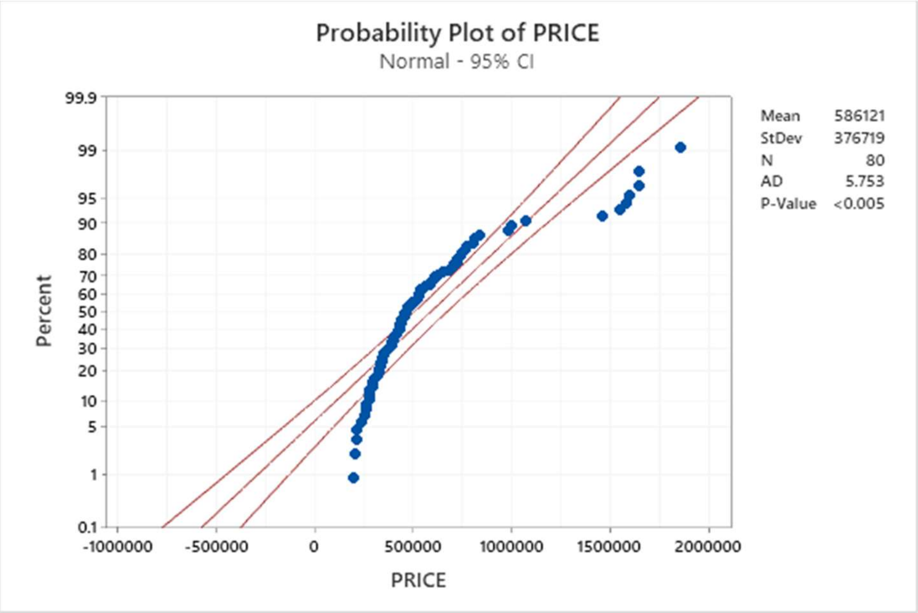


Figure 2: Probability Plot of Price (Y)

Since a more symmetric and normally distributed dependent variable is preferred for regression analysis, a transformation is performed on Price. We take the Log of Price to the base 10. The Table 2 shows the sample subset of 10 of the 80 properties with an added column showing the transformed variable Log (Y).

Table 2: Sample Subset of Dataset with Log of Price

	Y	Log (Y)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Property	PRICE	Log (Price)	bedrooms	bathrooms	sqft_living	sqft_lot	floors	Numbers of times viewed	Quality Grade	sqft_above	sqft_basement	Built or Renovated
1	$ 440,000.00	5.643452676	3	2.5	1910	66211	2	0	7	1910	0	1997
2	$ 213,000.00	5.328379603	2	1	1000	10200	1	0	6	1000	0	1961
3	$ 563,500.00	5.75089392	4	1.75	2085	174240	1	0	7	1610	475	1964
4	$ 1,550,000.00	6.190331698	5	4.25	6070	171626	2	0	12	6070	0	1999
5	$ 1,600,000.00	6.204119983	6	5	6050	230652	2	3	11	6050	0	2001
6	$ 350,000.00	5.544068044	3	2.25	1580	47916	1	0	7	1580	0	1979
7	$ 540,000.00	5.73239376	3	2.25	2000	217800	2	0	8	2000	0	1996
8	$ 535,000.00	5.728353782	3	1	1330	40259	1	0	7	1330	0	1977
9	$ 600,000.00	5.77815125	2	2.5	2410	102366	1	0	7	1940	470	1989
10	$ 275,000.00	5.439332694	3	1	1370	17859	1	0	7	1150	220	1930

From Figure 3, which depicts the histogram of the Log of Price it is observed that the graph is more symmetric as compared to Figure 1.

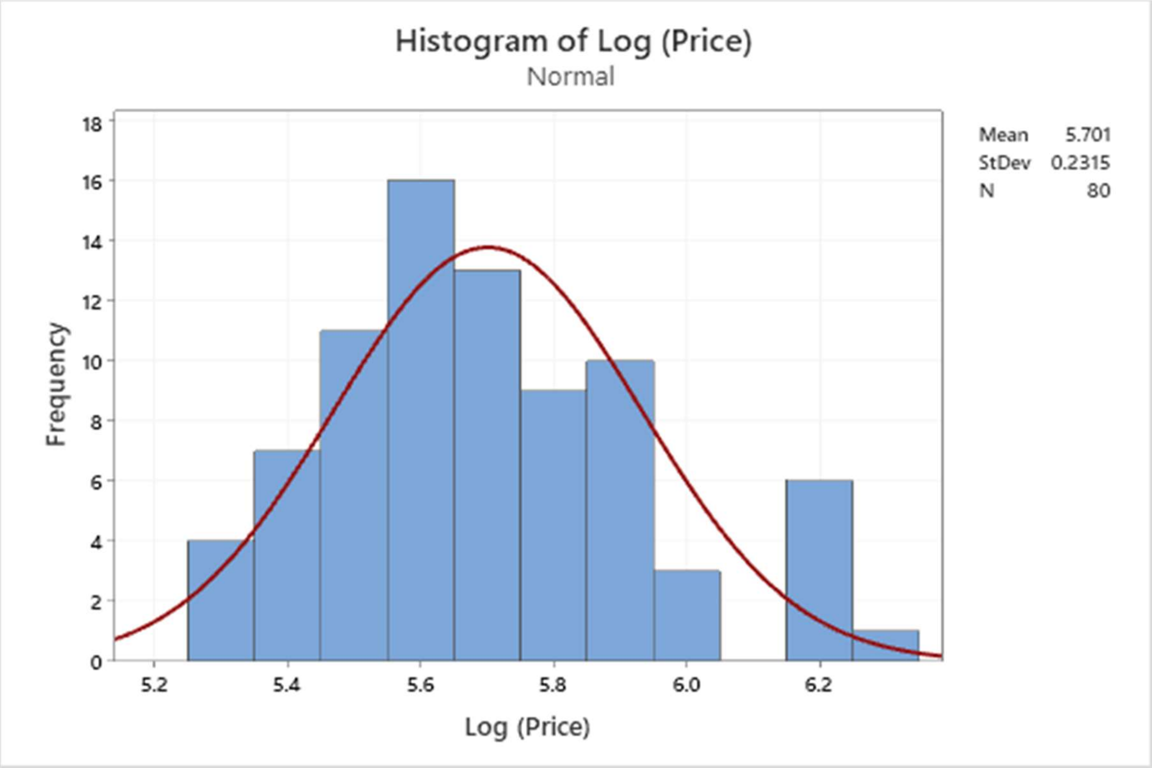


Figure 3: Histogram of Log Price (Log(Y))

Further, plotting the probability plot of Log (y) in Figure 4, it is observed that:

- Most of the points lie on a straight line
- Not many points lie outside the confidence boundaries of the plot
- There is less deviation from normality in Figure 4 than in Figure 2
- The p value if the plot is higher compared to the p value obtained in Figure 2
- The standard deviation of Log (Y) is much smaller compared to Y

Thus, Log (Y) – Log of Price is used as the dependent variable in the model.

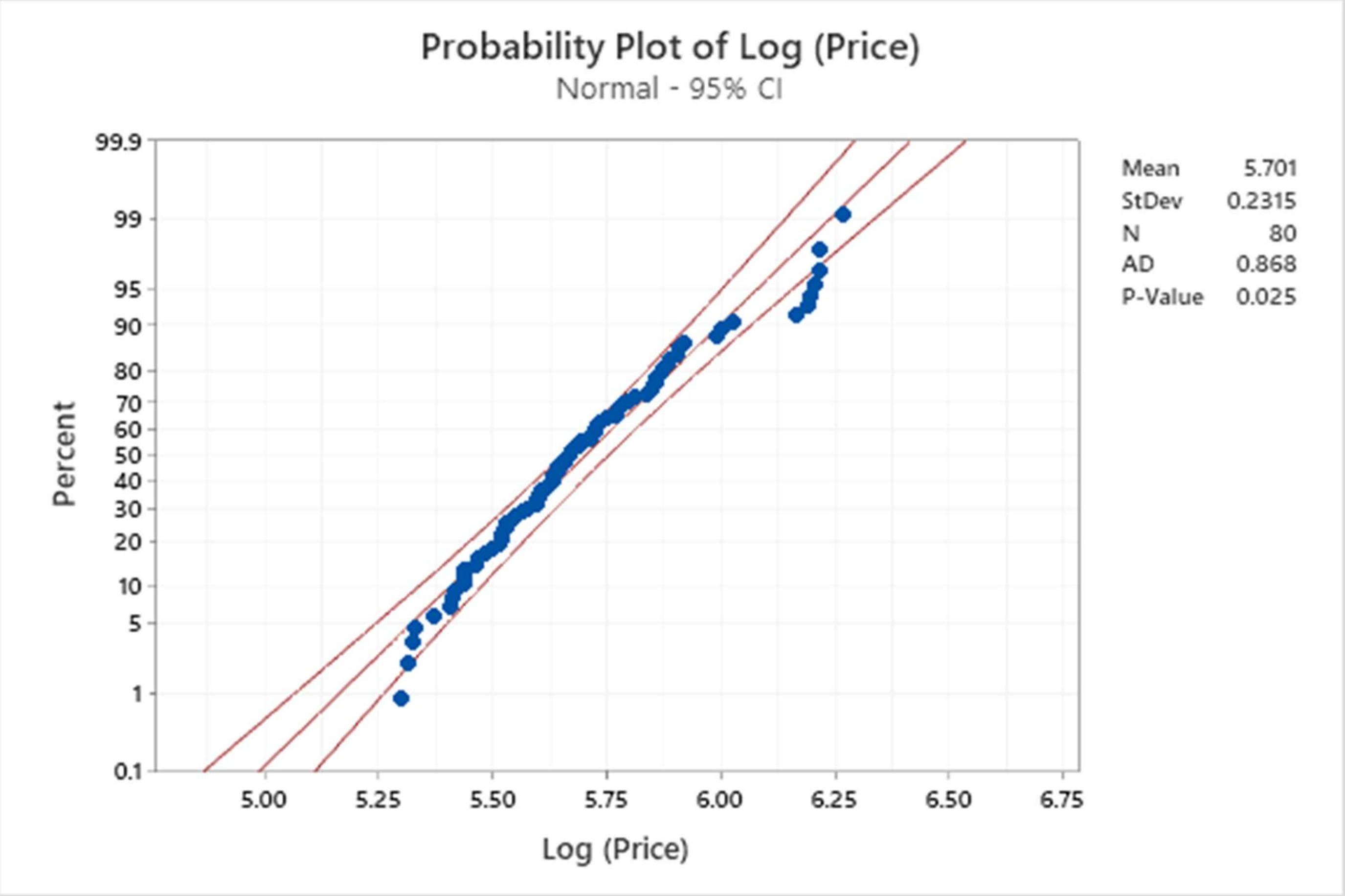


Figure 4: Probability Plot of Log Price (Log(Y))

To obtain a feel for the data the correlations between the dependent variable – Log of Price and the explanatory variable is studied. Correlation is a measure of linear dependence and thus is a good metric to understand the dependence of the variables for linear regression. Table 3 depicts the correlation matrix of the dataset. The second column of the table shows the correlation of our Log (Y) and the explanatory variables.

Higher correlation indicates larger linear dependence, thus it is reasonable to include explanatory variables with higher correlation in the model. To decide which variables to include, a threshold was chosen. The threshold chosen here to indicate significant correlations is **0.65**. Correlations with a value larger than the chosen threshold are highlighted in red in the table. The threshold was chosen such that a reasonable number out of the 10 explanatory variables could be highlighted as significant.

Immediately 5 variables stand out as significantly correlated to Log (Y):
X2 – bathrooms
X3 – sqft_living
X5 – floors
X7 – Quality Grade
X8 – sqft_above
X10 – Built or Renovated

A model could be built including all 5 variables, however what is also noticeable from the correlation matrix is that the variables X2, X3, X7 and X8 are all highly correlated to each other. Instead of including all of these variables in the initial model, using one of them may also be able to sufficiently explain the variation of all.

Table 3: Correlation Analysis of Log Price and Explanatory Variables Threshold: 0.65

	Log (Price)	*bedrooms*	*bathrooms*	*sqft_living*	*sqft_lot*	*floors*	*No. of times viewed*	*Quality Grade*	*sqft_above*	*sqft_basement*	*Built or Renovated*
	Log (Y)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Log (Y)	1.0000										
X1	0.5232	1.0000									
X2	0.8389	0.5591	1.0000								
X3	0.8933	0.6001	0.8894	1.0000							
X4	0.6162	0.1116	0.4296	0.4765	1.0000						
X5	0.6621	0.4145	0.6297	0.6252	0.4725	1.0000					
X6	0.4455	0.1968	0.4129	0.4015	0.6350	0.4192	1.0000				
X7	0.8876	0.4620	0.8302	0.8466	0.5389	0.6842	0.4147	1.0000			
X8	0.8854	0.5893	0.8582	0.9606	0.4503	0.6690	0.3766	0.8443	1.0000		
X9	0.0436	0.0487	0.1261	0.1574	0.1014	-0.1449	0.0953	0.0229	-0.1233	1.0000	
X10	0.6719	0.3131	0.7697	0.6445	0.3213	0.5520	0.2215	0.7010	0.6475	0.0005	1.0000

X2 shows significant correlation to → X3, X7, X8, and X10
X3 shows significant correlation to → X2, X7, and X8
X5 shows significant correlation to → X7, and X8
X7 shows significant correlation to → X2, X3, X5, X7, and X8
X8 shows significant correlation to → X2, X3, X5, X7, and X8
X10 shows significant correlation to → X2, and X7

Since the aim of regression is to build the most parsimonious model that would describe the variation in the variables, the variables **X3** – the variable with largest correlation to Log (Y), **X5** and **X10** are used to build the first model. These variables are highlighted in yellow in Table 3.

First Model

Regression Equation

Log (Price) = 3.01 + 0.000130 sqft_living + 0.0653 floors + 0.001158 Built or Renovated

Model Summary

S	R-sq	R-sq(adj)
0.0987939	82.47%	81.78%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	3.4905	1.16350	119.21	0.000
Error	76	0.7418	0.00976		
Total	79	4.2323			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.01	1.13	2.67	0.009	
sqft_living	0.000130	0.000012	10.44	0.000	2.08
floors	0.0653	0.0299	2.18	0.032	1.75
Built or Renovated	0.001158	0.000582	1.99	0.050	1.82

Durbin-Watson Statistic

Durbin-Watson Statistic = 1.89356

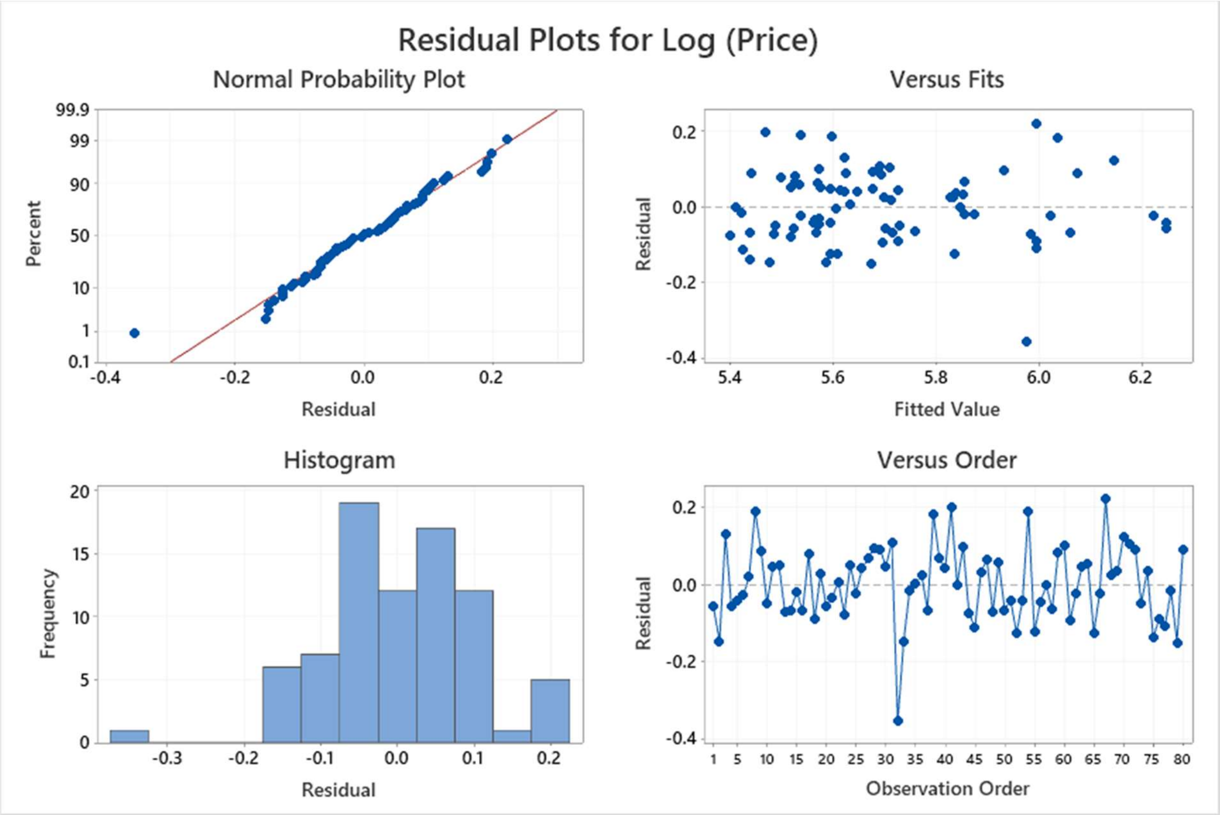


Figure 5: Residual Plots for Log (Y) for Model 1

Analysis for first model:

1. R – Squared = 82.47% and Adjusted R – Squared = 81.78%. Both the values are quite high and indicate this model is a good fit.
2. From Figure 6, we can observe that apart from one outlier most of the points lie on a straight line and within the confidence boundaries. The p – value of Residuals is also greater than 0.25. This indicates normal distribution thus we can perform statistical analysis on coefficients.
3. From Figure 6, Standard Deviation of Residuals = 0.09879 which is preferred since it is a low value.
4. The F-Value = 119.21 is significantly high.
5. The p – value of regression is 0, so we reject the Null Hypothesis that all coefficients are simultaneously = 0, indicating that at least 1 coefficient is not 0.
6. The p – value of individual variables is small, indicating that their respective coefficients is not 0 which is preferable.
7. The VIF for all the variables are over 1 indicating that some collinearity, but under 5 which means that the regression coefficients are NOT poorly estimated.
8. The Durbin-Watson Statistic is 1.89, falling in the range of 1.5 – 2.5, indicating little to no autocorrelation in the residuals.
9. The Residual VS Fits graph from Figure 5 seems to be evenly spaced and doesn't depict any patterns, showing the residuals have constant variation and no Heteroscedasticity.
10. The Residual VS Order graph from Figure 5 appears to be chaotic indicating little autocorrelation.
11. Overall, the model is good enough, but it can be improved by adding explanatory variables to the model to increase R – Squared Value
12. The next model will be evaluated using the extra variable *Quality Grade* since it is the most correlated to Log (Y) among the variables highlighted from Table 3.

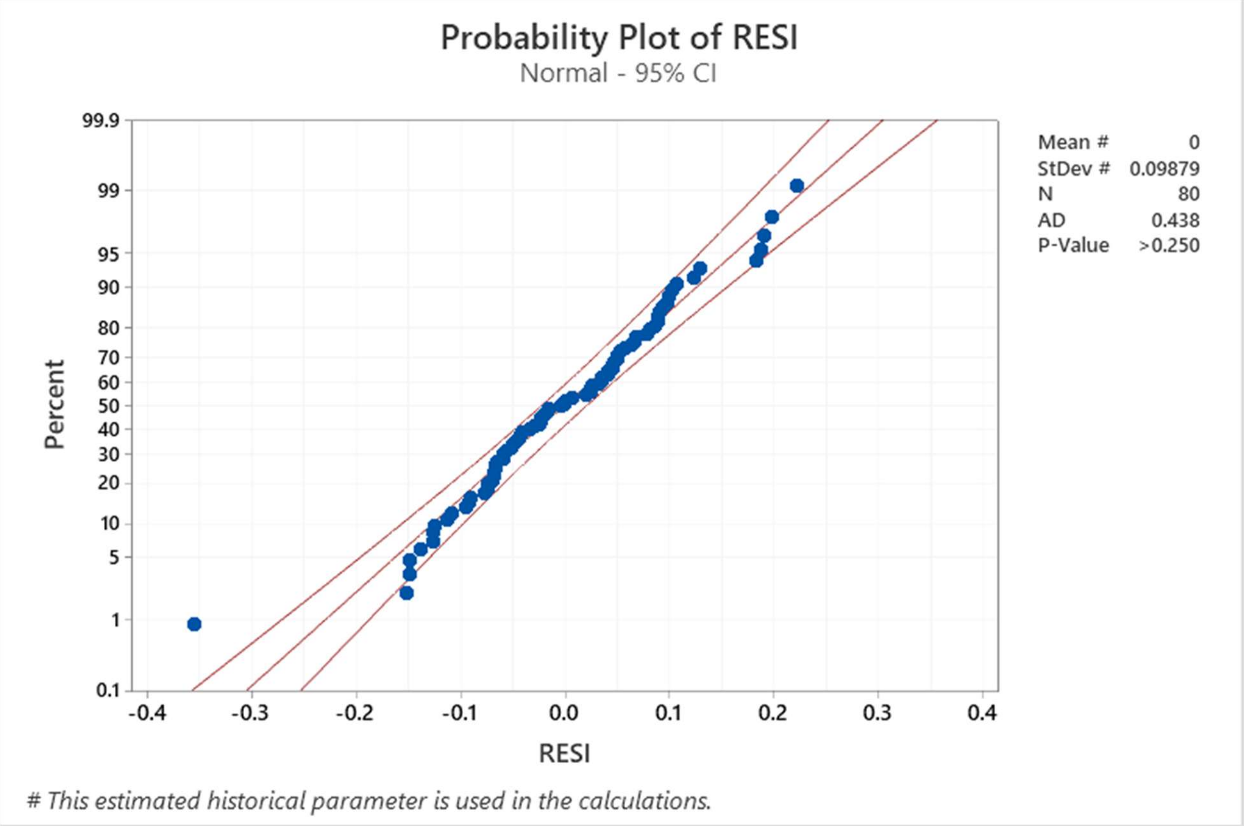


Figure 6: Probability Plot of Residual for Model 1

Second Model

Regression Equation

Log (Price) = 4.25 + 0.000087 sqft_living + 0.0262 floors + 0.000369 Built or Renovated + 0.0628 Quality Grade

Model Summary

S	R-sq	R-sq(adj)
0.0883693	86.16%	85.42%

Analysis of Variance

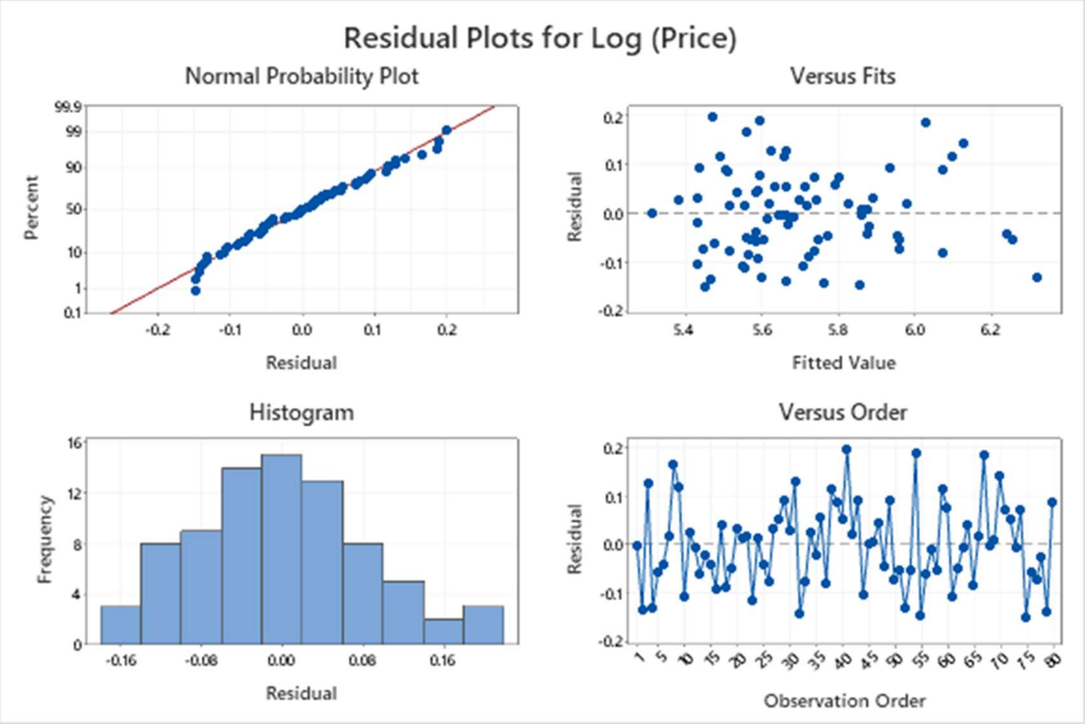
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	3.6466	0.911646	116.74	0.000
Error	75	0.5857	0.007809		
Total	79	4.2323			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.25	1.05	4.07	0.000	
sqft_living	0.000087	0.000015	5.92	0.000	3.63
floors	0.0262	0.0282	0.93	0.355	1.94
Built or Renovated	0.000369	0.000550	0.67	0.505	2.03
Quality Grade	0.0628	0.0140	4.47	0.000	4.57

Durbin-Watson Statistic

Durbin-Watson Statistic = 2.07784


Residual Plots for Log (Price)

Normal Probability Plot

Versus Fits

Histogram

Versus Order

Figure 7: Residual Plots for Log (Y) for Model 2

Analysis for second model:

1. R – Squared = 86.16% and Adjusted R – Squared = 85.42%. Both the values are high and have improved from the previous model indicating this model is a good fit.
2. From Figure 8, we can observe that apart from one outlier most of the points lie on a straight line and within the confidence boundaries. The p – value of Residuals is also greater than 0.25. This indicates normal distribution thus we can perform statistical analysis on coefficients.
3. From Figure 8, Standard Deviation of Residuals = 0.08837 which is lower than model 1.
4. The F-Value = 116.74 is significantly high.
5. The p – value of regression is 0, so we reject the Null Hypothesis that all coefficients are simultaneously = 0, indicating that at least 1 coefficient is not 0.
6. The p – value of *sqft_living* and *Quality Grade* is small, indicating that their respective coefficients is not 0 which is preferable. However, the p – values for *floors* = 0.355 (35.5%) and *Built or Renovated* = 0.505 (50.5 5) has significantly increased from Model 1 which is not ideal.
7. The VIF for all the variables are over 1 indicating that some collinearity, but under 5 which means that the regression coefficients are NOT poorly estimated.
8. The Durbin-Watson Statistic is 2.07784, falling in the range of 1.5 – 2.5, indicating little to no autocorrelation in the residuals.
9. The Residual VS Fits graph from Figure 7 seems to be evenly spaced and doesn't depict any patterns, showing the residuals have constant variation and no Heteroscedasticity.
10. The Residual VS Order graph from Figure 7 appears to be chaotic indicating little autocorrelation.
11. Overall, the model is good enough, but it can be improved by dropping explanatory variables with high p - values
12. The next model will be evaluated without the extra variable *Built or Renovated* since it has the highest p – value.

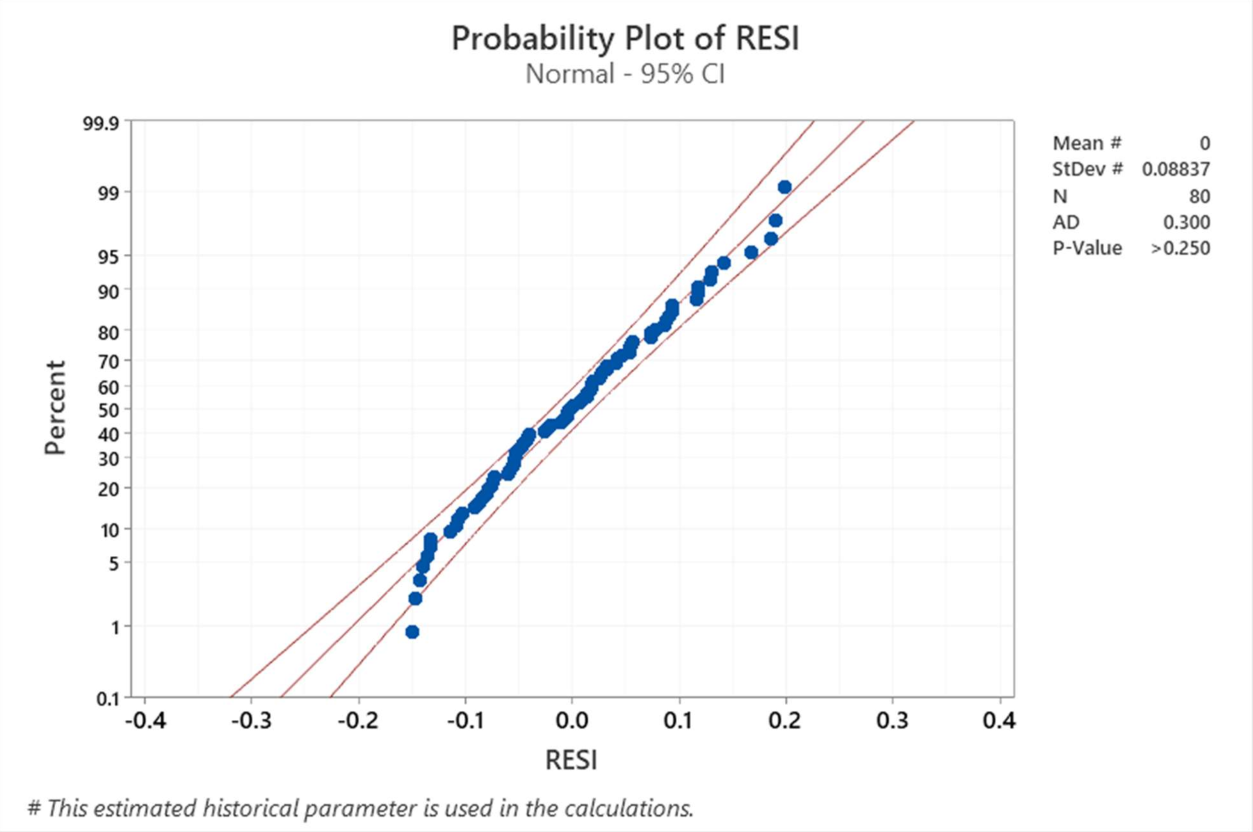


Figure 8: Probability Plot for Residual for Model 2

Third Model

Regression Equation

Log (Price) = 4.9503 + 0.000088 sqft_living + 0.0658 Quality Grade + 0.0286 floors

Model Summary

S	R-sq	R-sq(adj)
0.0880488	86.08%	85.53%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	3.6431	1.21436	156.64	0.000
Error	76	0.5892	0.00775		
Total	79	4.2323			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.9503	0.0679	72.89	0.000	
sqft_living	0.000088	0.000015	6.06	0.000	3.58
Quality Grade	0.0658	0.0133	4.97	0.000	4.10
floors	0.0286	0.0279	1.03	0.308	1.91

Durbin-Watson Statistic

Durbin-Watson Statistic = 2.10314

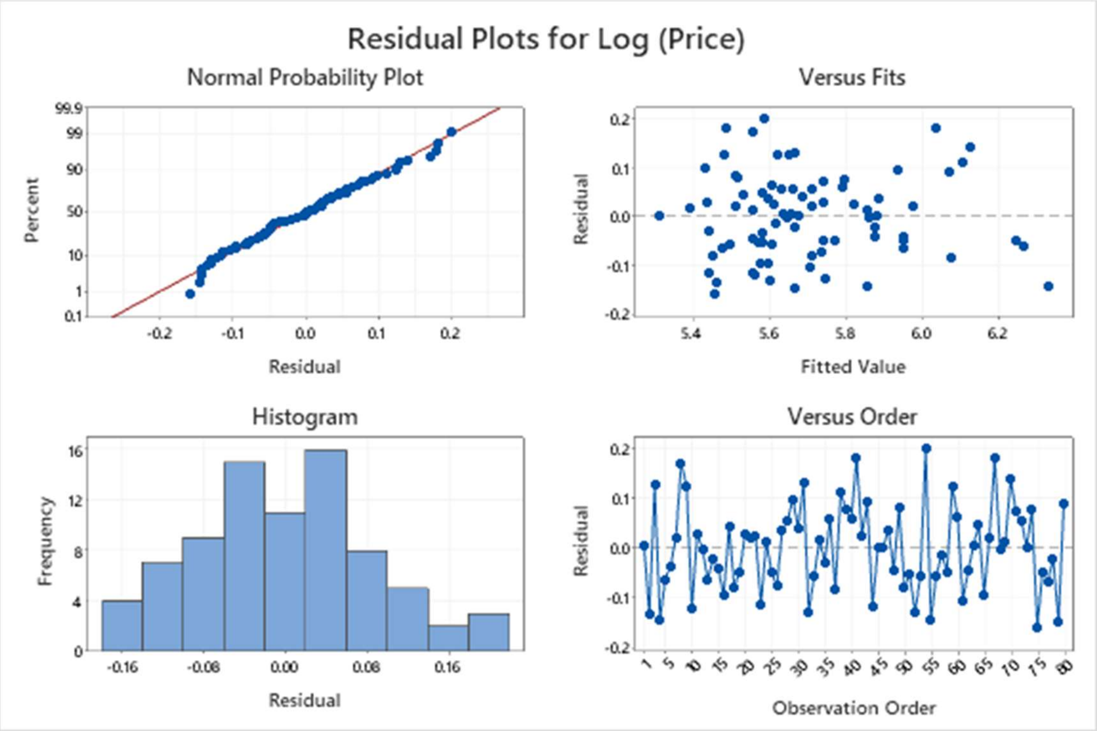


Figure 9: Residual Plots for Log (Y) for Model 3

Analysis for third model:

1. R – Squared = 86.08% and Adjusted R – Squared = 85.53%. R – Squared has dropped due to dropping of a variable, but the value of Adjusted R – Squared has still increased. Both values are significantly high enough to indicate a good fit.
2. From Figure 10, we can observe that apart from one outlier most of the points lie on a straight line and within the confidence boundaries. The p – value of Residuals is also greater than 0.25. This indicates normal distribution thus we can perform statistical analysis on coefficients.
3. From Figure 10, Standard Deviation of Residuals = 0.08805 which is lower than model 2.
4. The F-Value = 156.64 is significantly high.
5. The p – value of regression is 0, so we reject the Null Hypothesis that all coefficients are simultaneously = 0, indicating that at least 1 coefficient is not 0.
6. The p – value of *sqft_living* and *Quality Grade* is small, indicating that their respective coefficients is not 0 which is preferable. The p – values for *floors* = 0.308 (30.8%) which is not ideal.
7. The VIF for all the variables are over 1 indicating that some collinearity, but under 5 which means that the regression coefficients are NOT poorly estimated.
8. The Durbin-Watson Statistic is 2.10314, falling in the range of 1.5 – 2.5, indicating little to no autocorrelation in the residuals.
9. The Residual VS Fits graph from Figure 9 seems to be evenly spaced and doesn't depict any patterns, showing the residuals have constant variation and no Heteroscedasticity.
10. The Residual VS Order graph from Figure 9 appears to be chaotic indicating little autocorrelation.
11. At this point, we have a good model with high R- Squared an Adjusted R – Squared Values, low p- values on individual variables in general, VIF values below 5 for all variables and no signs of auto correlation and heteroscedasticity.
12. We can test if this model can be improved by exploring the relation and interaction of explanatory variables and adding an interaction variable to the model.
13. We plot the relation of various explanatory variables with Log (Y) to see if any of the plots stand out. One plot, depicting interaction between *sqft_basement* and *bedrooms* in Figure 11 showed significant variation and thus an interaction term [**sqft_basement * bedrooms**], was added to capture their interaction in a model.

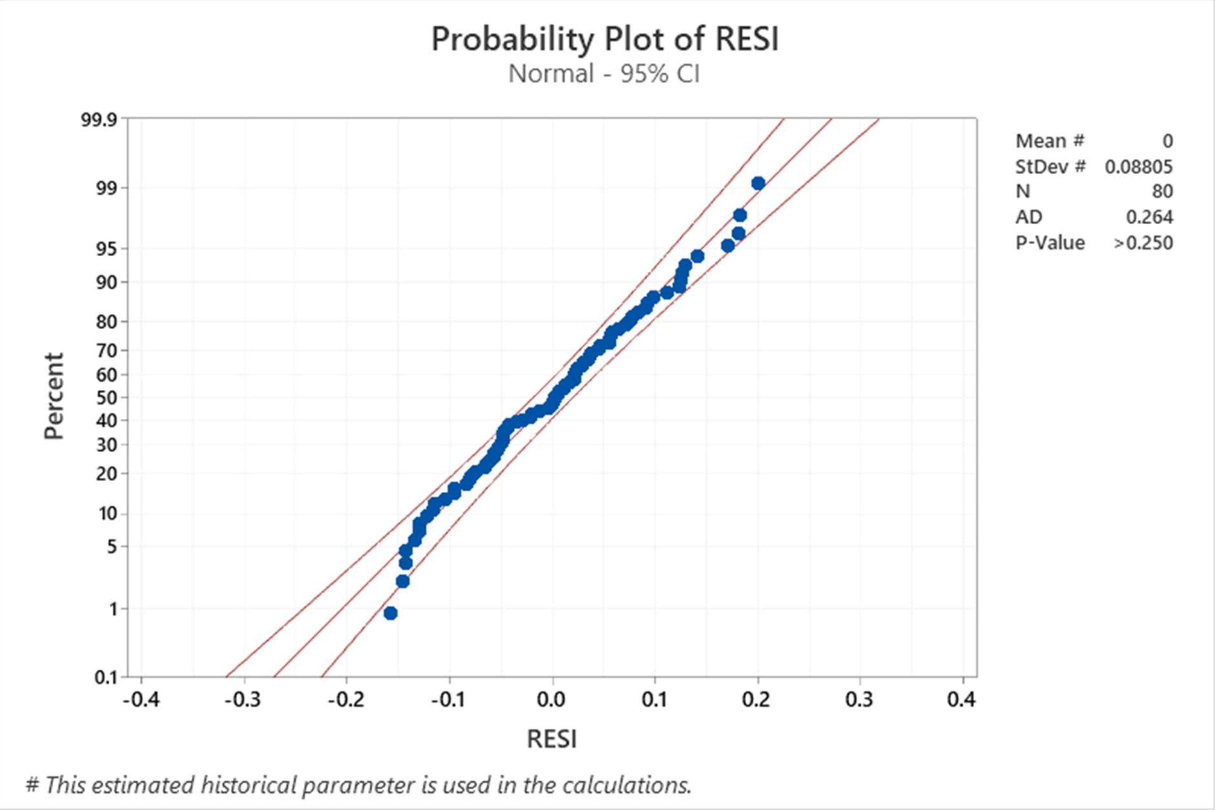


Figure 10: Probability Plot for Residual for Model 3

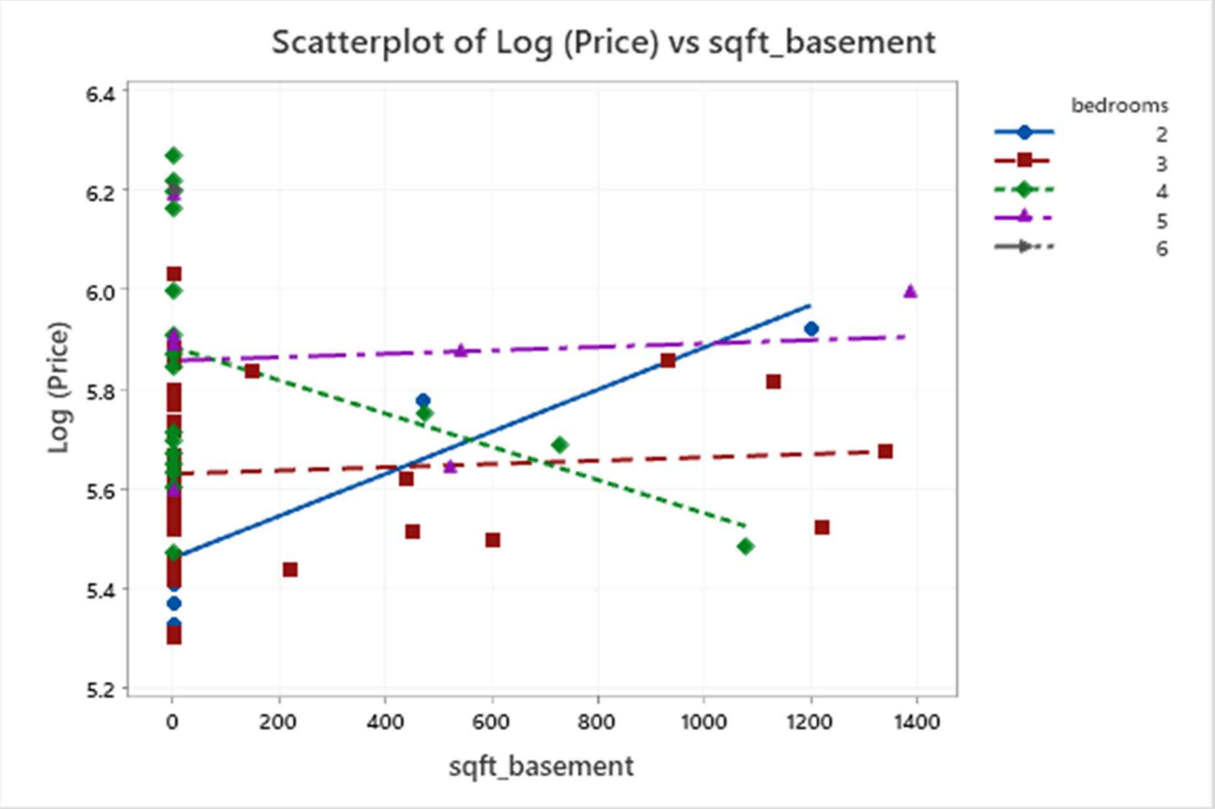


Figure 11: Interaction Between sqft_basement and bedrooms

Fourth Model

Regression Equation

Log (Price) = 4.9520 + 0.000077 sqft_living + 0.0268 floors + 0.0629 Quality Grade + 0.000077 Sqft above / bedrooms

Model Summary

S	R-sq	R-sq(adj)
0.0877423	86.36%	85.63%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	3.6549	0.913716	118.68	0.000
Error	75	0.5774	0.007699		
Total	79	4.2323			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.9520	0.0677	73.15	0.000	
sqft_living	0.000077	0.000017	4.49	0.000	5.00
floors	0.0268	0.0278	0.97	0.337	1.91
Quality Grade	0.0629	0.0134	4.68	0.000	4.23
Sqft above / bedrooms	0.000077	0.000062	1.24	0.220	3.44

Durbin-Watson Statistic

Durbin-Watson Statistic = 2.05878

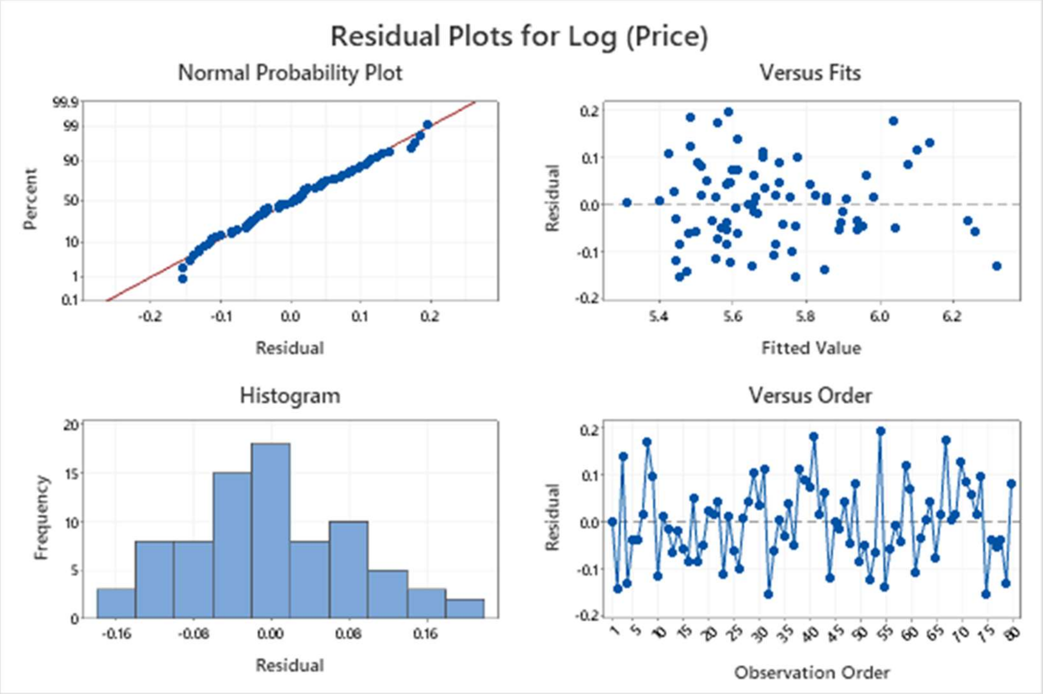


Figure 12: Residual Plots for Log (Y) for Model 4

Analysis for fourth model:

1. R – Squared = 86.36% and Adjusted R – Squared = 85.63%. R – Squared has increased due to adding of a variable, but the value of Adjusted R – Squared has also increased. Both values are significantly high enough to indicate a good fit.
2. From Figure 13, we can observe that apart from one outlier most of the points lie on a straight line and within the confidence boundaries. The p – value of Residuals is also greater than 0.25. This indicates normal distribution thus we can perform statistical analysis on coefficients.
3. From Figure 12, Standard Deviation of Residuals = 0.08774 which is lower than model 3.
4. The F-Value = 11.868 is significantly high but lower than model 3.
5. The p – value of regression is 0, so we reject the Null Hypothesis that all coefficients are simultaneously = 0, indicating that at least 1 coefficient is not 0.
6. The p – value of *sqft_living* and *Quality Grade* is small, indicating that their respective coefficients is not 0 which is preferable. The p – values for *floors* = 0.337 (33.7%) which is not ideal.
7. The VIF for all the variables except *sqft_living* are over 1 indicating some collinearity, but under 5 which means that the regression coefficients are NOT poorly estimated. *sqft_living* in this model may be poorly estimated. While the increase in VIF is not ideal, it is expected because we added an
8. The Durbin-Watson Statistic is 2.05878, falling in the range of 1.5 – 2.5, indicating little to no autocorrelation in the residuals.
9. The Residual VS Fits graph from Figure 12 seems to be evenly spaced and doesn't depict any patterns, showing the residuals have constant variation and no Heteroscedasticity.
10. The Residual VS Order graph from Figure 12 appears to be chaotic indicating little autocorrelation.
11. At this point, both Models 3 and 4 have high R- Squared an Adjusted R – Squared Values, low p- values on individual variables in general, VIF values below 5 for - all variables (Model 3) and all but one variable (Model 4), and no signs of auto correlation and heteroscedasticity.
12. While the 4th model has higher values R – Square and Adjusted R – Square, we must evaluate if this increase is statistically significant enough to warrant adding another variable to the model. We do this using an F – hypothesis test.

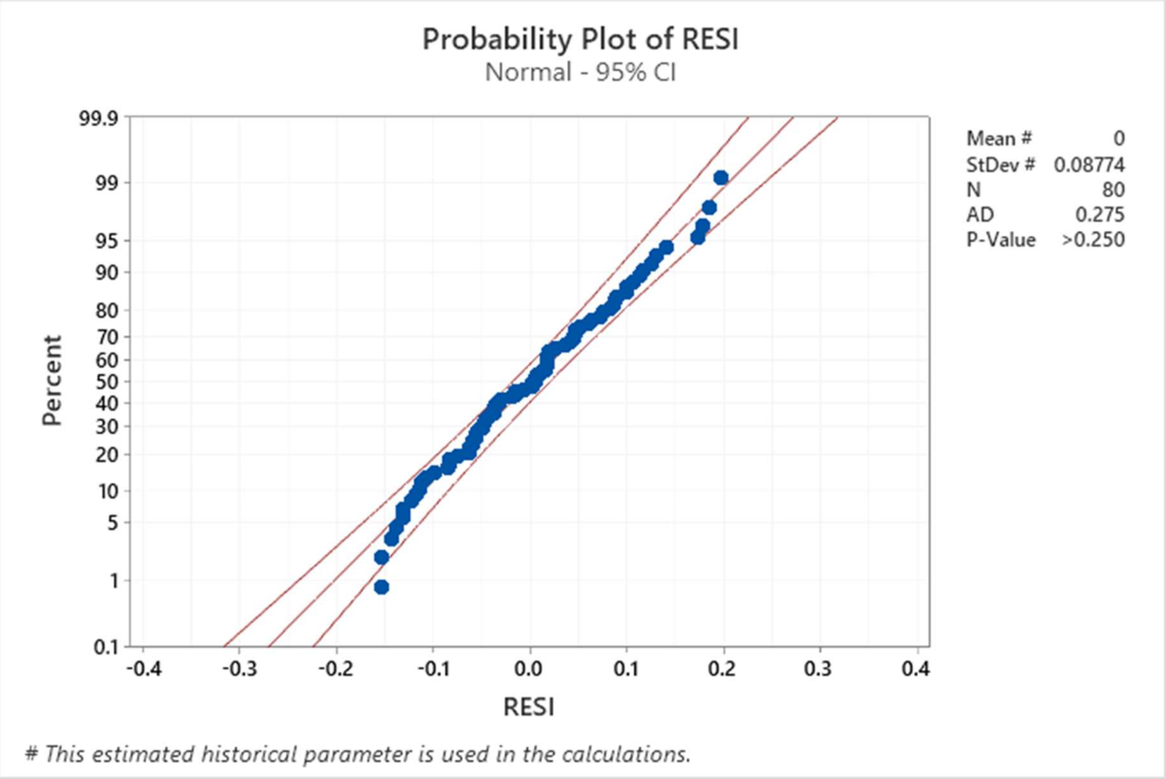


Figure 13: Probability Plot for Residual for Model 4

Since all the variables of Model 4 are present in Model 3, we can perform the increase in R² test.

H0: No Model Improvement

H1: Model Improvement

We calculate the F – statistic as:

$$F = \frac{(R_4^2 - R_3^2)/\,(df_3 - df_4)}{(1 - R_4^2)/df_4}$$

Model 4	
R Square	86.36%
Degrees of Freedom (df)	75
Model 3	
R Square	86.08%
Degrees of Freedom	76

Explanatory Variables in the model 4

floors	Sqft_living	quality grade	sqft / bedrooms

Explanatory Variables in the model 3

floors	Sqft_living	quality grade

Conclusion: All variables of small/restricted model are variables in the full model and "the increase in R² test" can be performed

Difference R-Squared	0.28%
Difference df	1

	Value		
Numerator	0.0028		
Denominator	0.0018		
F-Statistic	1.540		
⍰	5%		
Critical Value	3.968	**Conclusion**	**Method 1**
Conclusion	Fail to Reject H0	No Model Improvement	
p-value	21.85%	**Conclusion**	**Method 2**
Conclusion	Fail to Reject H0	No Model Improvement	

H_0: No Model Improvement

H_1: Model Improvement

As we can see from the F- hypothesis test, adding the interaction term variable did not make any significant statistical changes in R – squared value. Thus adding another variable to the model is not warranted when the aim of linear regression is to make the simplest model that can sufficiently describe the variation of the explanatory variables.

Thus Model 3 is chosen as the final model.

Log (Price) = 4.9503 + 0.000088 sqft_living + 0.0658 Quality Grade + 0.0286 floors

We cannot determine if an observation is influential simply by checking if it is an outlier. Some outliers may be difficult to detect from residuals. Behavior of residuals don't determine all influential observations.

The DFIT coefficient of each observation gives the combined effect of a single observation on all regression coefficients. Any |DFIT| observation that is greater than the threshold is considered influential.

Influential observations can also be calculated using Studentized Residuals.
The Table 4 shows influential Studentized Residuals (TRES) and DFITS of the observations in the Final Model.

p - Parameters	3
n - Observations	80
DFIT THRESHOLD	**0.447214**

$$DFIT\ Threshold = 2\sqrt{(p+1)/n}$$

a	10%
n-p-2	75
TRES Threshold	1.665425

TRES Threshold = 95th percentile of Student – T dist.

$$TRES1\ Threshold = T_{n-p-2}$$

Table 4: TRES and DFIT values of influential observations

Data	TRES	DFIT
4	-1.78578	-0.73781
8	2.014012	0.356765
32	-1.75148	-1.09769
38	1.396144	0.640574
41	2.169496	0.478942
54	2.366732	0.371069
55	-1.68024	-0.33356
67	2.14923	0.463931
70	1.688583	0.499134
75	-1.84483	-0.3173
79	-1.71303	-0.3364

Prediction

We are given the following values and we need to predict Y

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Property	PRICE	bedrooms	bathrooms	sqft_living	sqft_lot	floors	Numbers of times viewed	Quality Grade	sqft_above	sqft_basement	Built or Renovated
Forecast	????	3	2	2250	50000	2	0	6	750	0	2000

We can't directly predict for Y, so first prediction for Log (Y) is evaluated.

Regression Equation

Log (Price) = 4.9503 + 0.000088 sqft_living + 0.0658 Quality Grade + 0.0286 floors

Settings

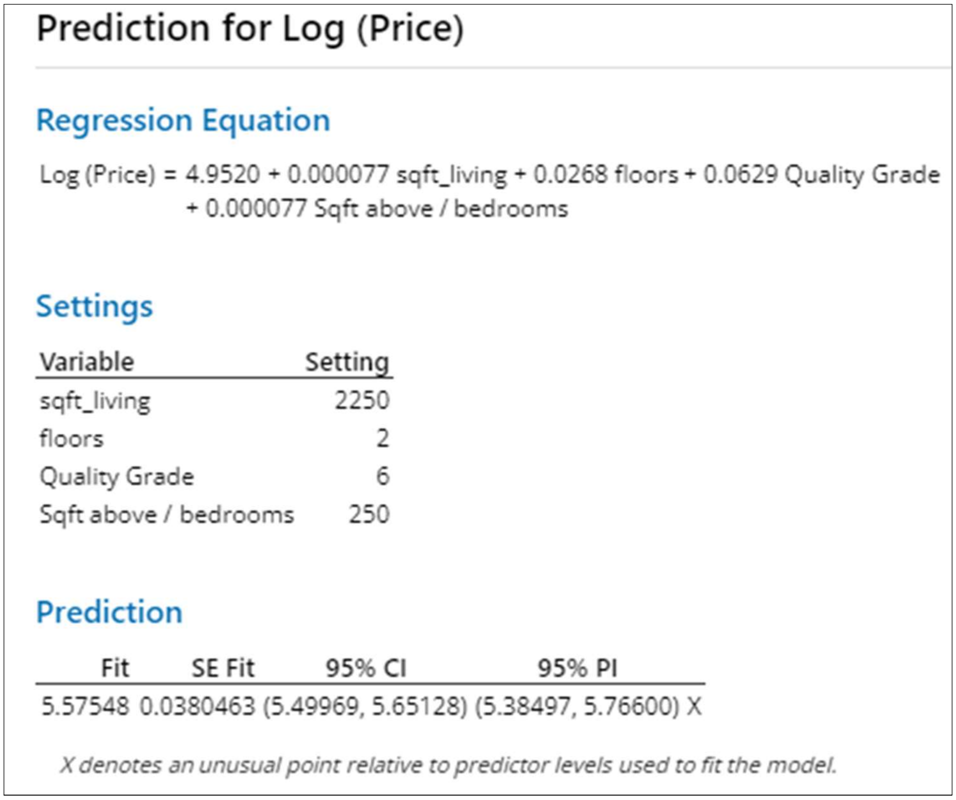
Variable	Setting
sqft_living	2250
Quality Grade	6
floors	2

Prediction

Fit	SE Fit	95% CI	95% PI
5.60121	0.0319785	(5.53752, 5.66490)	(5.41464, 5.78778)

Prediction Interval Width Model 3 = 5.78778 - 5.41464 = 0.37314

In Comparison, for Model 4:



Prediction for Log (Price)

Regression Equation

Log (Price) = 4.9520 + 0.000077 sqft_living + 0.0268 floors + 0.0629 Quality Grade
+ 0.000077 Sqft above / bedrooms

Settings

Variable	Setting
sqft_living	2250
floors	2
Quality Grade	6
Sqft above / bedrooms	250

Prediction

Fit	SE Fit	95% CI	95% PI	
5.57548	0.0380463	(5.49969, 5.65128)	(5.38497, 5.76600)	X

X denotes an unusual point relative to predictor levels used to fit the model.

Prediction Interval Width Model 4 = 5.766 - 5.38497 = 0.38103

Since the Prediction Interval width for Model 3 is smaller than Model 4, it is most ideal to go ahead with Model 3.

The credibility interval gives the uncertainty of the prediction. We can see from the below Prediction Analysis that the Credibility Interval is too large leaving a lot of space for uncertainty in the analysis.

PFITS	PSEFITS	CLIM	CLIM_1	PLIM	PLIM_1
5.601209581	0.03197847	5.537518944	5.664900218	5.41463745	5.787781716

m = LOG(PRICE) - hat	5.601	5.600
MEDIAN[PRICE]	$399,217.51	
E[PRICE]	$408,613.45	

		Variances
Standard Error Residuals	0.088050	0.007753
Standard Error LOG(Price-hat)	0.031978	0.001023
s^2 = Var[Y]=Var[Log(Price)]		0.008775
s = Standard Deviation [Log(Price)]	0.093677	

95% Confidence Interval

LB E[LOG(PRICE)]	5.53752
UB E[LOG(PRICE)]	5.66490

95% Prediction Interval (or Credibility Interval)

LB LOG(PRICE)	5.414637
UB LOG(PRICE)	5.787782

Approximate 95% Confidence Interval

LB E[PRICE]	$344,761.64
UB E[PRICE]	$462,274.80

95% Prediction Interval (or Credibility Interval)

PRICE	$259,799
PRICE	$613,453.6

Approximate because Log is not a linear function

Conclusion: Although Model 3 was decided as the most apt model to predict with, evaluating the credibility interval still leaves a lot of space for uncertainty in the values. The difference between the UB and LB of the Credibility Interval is $353,654.6 which would not be a desirable model to work with in real estate properties.