

To City of London Officials,

Although the widening of Springbank Drive will provide many benefits to our growing community, many homeowners will see their property values negatively affected due to increased traffic, front yard setbacks, and other problems that will negatively affect their values. Due to the Expropriations Act, homeowners are entitled to compensation. As such, it is of utmost importance to compensate homeowners appropriately, as not doing so would leave the city vulnerable to greater compensation claims.

While traditional housing appraisal relies on the subjective assessment of an appraiser, our valuation tool uses multiple linear regression models to objectively assess the appropriate price of a home. If the City of London were legally challenged over the compensation claim, our model provides a strong justification for why a homeowner shouldn't be compensated further. We look forward to showing you how our superior model will undoubtedly help save the City of London control the costs of the Springbank Drive project, so you can better serve the growing community.

Descriptive Statistics

To generate our models, we utilized a sample gathered from 104 homes that were located around Springbank Drive and sold from January 1998 to May 2003. From these records, we generated the following descriptive statistics.

Table 1: Descriptive Statistics of 104 Homes around Springbank Drive

	Units	Mean	Standard Deviation	Median	Max	Min
Home Price	\$	134,469	27,744	131,400	249,900	69,900
Four Lane Road ^{1*}	-	0.33	0.47	0	1	0
Traffic Count (Cars) ²	Count	20,221	8,369	19,000	36,000	0
Lot Frontage	Sq. Ft.	902.85	255.55	863	1,674	0
Home Age	Years	49.52	15.7	48	118	0
Basement Finished Area	-	396.78	287.16	471	1,134	0
Stories*						
One and a Half	-	0.09	0.28	0	1	0
Two	-	0.15	0.36	0	1	0
Interior Condition*						
Average	-	0.45	0.5	0	1	0
Good	-	0.34	0.47	0	1	0
Excellent	-	0.12	0.32	0	1	0
View*						
Average	-	0.58	0.5	1	1	0
Good	-	0.22	0.42	0	1	0
Exterior Amenities*						
Minor	-	0.22	0.42	0	1	0
Two or Three	-	0.42	0.5	0	1	0
More Than Three	-	0.13	0.34	0	1	0

1 = Only Included in Model 1

2 = Only Included in Model 2

** = Binary Variable. Either Applicable or Not Applicable to a given property.*

Although variables such as Lot Frontage, Traffic Count, Home Price and Basement Finished Area have the largest numbers relative to the other variables across all descriptive statistics, the scale of measurements used for each of these variables is dramatically different than for the other variables, particularly the binary variables. As such, it is not useful to compare variables between each other based on the descriptive statistics used.

The binary variables give the most fruitful data, as the arithmetic means translate to a percent value. For example, the average view of 0.58 means that 58% of homes have an average view. In addition, the dummy variables that are binary in nature are useful for representing qualitative measures like the view or condition of the home. These variables are important as it pertains to home valuation, so their inclusion is imperative.

Methodology and Model

In contrast with traditional sales appraising, our multiple regression centered methodology is a more accurate measure of real estate valuation for the following reasons. The first reason is that our model pertains to the issue at hand of the widening of Springbank Drive. Oftentimes, things like the amount of traffic are not necessarily the “go to” comps for traditional real estate. However, in this context, the traffic amount is very relevant as it pertains to compensation from the widening of the road. In addition, real estate comps can often be subjective, with very little overlap regarding the similarity of variables from one home to another. Moreover, multiple regression attempts to value each additional variable holding all others constant, that is, if a home lacks certain variables, the value of what it does have is less skewed.

The variables included were chosen to represent factors that most explain the value of any respective home located around Springbank Drive. Factors including a home’s size, condition, number of amenities, view, etc. are often the biggest drivers in price. Ordinarily, location would be the most significant factor, however, all these homes are intentionally located near each other, so this wasn’t included. Multiple regression requires that each variable, and the associated coefficient make intuitive sense, for example, depending on the model, each additional SF of finished basement adds between \$21.08 - \$23.96 to the value of the home.

Moreover, because we know we can satisfy the assumption of normality, the methodology of using multiple regression to find compensation figures is sounder in its approach. By viewing the standardized residual plots (see final page), it’s clear that the most all the estimates for price fall within two standard deviations of the mean of home price, as well the model doesn’t tend to either overpredict or underpredict homes prices, that is the residuals are normally distributed around zero.

Variables

Models were developed based on the variables in the following table. These models were used as regressors to predict the appropriate compensation of the five homes impacted by the widening of Springbank drive.

Table 2 - Variables

#	Variable	Justification for Inclusion in Models
1	Traffic Count	Springbank Drive widening will increase traffic, which negatively affects real estate valuations.
2	Stories	This is a proxy for size, larger homes more often than not are more expensive than smaller homes.
3	Age	Age can be a reflection of wear and tear, ie homes that need immediate repairs will often have reduced final sales price
4	Amenities	The more amenities a home has, the more desirable it is
4	View	The better a view a home has, the more desirable it is
5	Interior Condition	The better a home’s condition, the more desirable it is

Results

Table 4 - Regression Results
Dependent Variable: Estimated Sale Price (in Canadian Dollars)

	Model 1		Model 2	
	<i>Traffic Count Model</i>		<i>Four Lane Rd. Model</i>	
	Estimate	Std. Error	Estimate	Std. Error
Intercept	80830.51***	21245.45	63927.39**	20393.97
One and a Half Stories	898.93	10828.89	988.88	11118.53
Two Stories	7996.15	7881.74	5700.37	8110.57
Home Age Years	168.83	204.13	137.52	210.90
Area Of Frontage	38.55***	10.77	40.68***	11.06
Minor Exterior Amenities	-4604.69	7530.08	-1380.35	7572.77
Two or Three Extra Amenities	-9482.84	7150.94	-8604.50	7359.07
More than Three Exterior Amenities	-8381.08	9417.20	-8900.82*	9781.14
Good View	12920.79	8580.98	17255.41*	8601.49
Average View	10634.56	6481.50	12017.45	6655.97
Average Interior Condition	9140.04	8782.19	10384.70	9036.36
Good Interior Condition	19856.64*	9308.55	20421.10*	9547.17
Excellent Interior Condition	25624.89*	11922.03	26964.08*	12215.55
Basement Finished Area	23.96*	10.33	21.08	10.87
Four Lane Road	-	-	-7966.89	6232.16
Traffic Count	-0.84*	0.33	-	-
R-squared	0.405		0.374	
Adjusted R-squared	0.312		0.275	

Discussion

There are few key indicators from the output above that suggest both models have validity, starting with the coefficients for Traffic Count and Four Lane Road. A coefficient of $-.84$ for traffic count suggests that holding all other variables constant, the value of each additional car that drives on the road in which the home in question is located, reduces the value of the home by \$0.84. In the case of Springbank Drive, it is expected from the previous case that this small coefficient makes sense, because there can be thousands of cars that pass by, and so the coefficient for each incremental car would be tiny, however when summed up, would make a significant impact on the price of the home. The same can be said for the variable Four Lane Road, however unlike the coefficient for Traffic Count, is rather one large negative number of $-\$7,967$ in reduction of value for a home that is located on a Four lane Road, holding all other variables constant. This means that, theoretically speaking, the exact same home located on a two-lane road would be worth \$7,967 more than if it were located on a Four Lane Road. Moreover, the coefficients for Area of Frontage make sense because holding all other variables constant, in model 1, each square foot of Lot frontage adds \$38.55, and in Model 2, adds \$40.68 to the value of a home. In the case of Springbank drive, these values are important because the city is reducing the Area of Frontage for 5 homes, meaning depending on the model, they should be compensated between the above \$38.55 and \$40.68 for each square foot of Area Frontage taken from them. Moreover, factors like interior condition make sense intuitively as each incremental increase in quality from Average, Good, and Excellent correspond with incremental increase in home prices, of \$9,140, \$19,857, and \$25,624 in Model 1 as well as \$10,385, \$20,421, and \$26,964 respectively. Broadly speaking, smaller coefficients (less than $|\$200|$) in the model is reflective of the incremental value with a one unit change in the associated variable, whereas larger coefficients translate to the value associated if a certain condition is met or not.

It's worth noting a few variables that may not be truly reflective of the appropriate associated values. The first being, the age of the home. In the models above, newer homes are worth between \$137 and \$168 dollars more for every year they are closer to the present. Although it is true that newer homes are less likely to need repairs, oftentimes older homes can be worth more, for a variety of reasons including types of homes, "Historic" homes/structures, etc.

Moreover, as far as the coefficients for the specific variables pertaining to Four Lane Road, as well as Traffic Count, there's always the possibility of correlation vs causation, that is, more expensive homes could tend to be built on smaller streets, which mean smaller traffic counts, which is reflected in the associated value of the coefficients in the above models. The same could be said with for the variables associated with 'view', that is, a property with a desirable view almost certainly won't have a low value home built on it; an individual who will pay a premium for a desirable lot because of a nice view, will almost certainly build a higher end home on the lot.

In addition, there could be some multi-collinearity with variables mentioned above, that is, the relationship with traffic count or four lane road and the associated measure of good vs average view. It's not a stretch to suggest that a home located on a four-lane road or with heavy traffic count is more likely to have an average view rather than a good view.

Implications

There are two possible formulas for the city to consider for final compensation figures for the impacted homes.

Model 1: Traffic Count Model

Compensation = \$12,000 (*Lump Sum*) + \$38.55 × (Sq. Ft. of Frontage Lost) + \$7,967 (*If home now on Four Lane Road*)

Property #	Lump Sum	L.F.	L.F. Comp.	4 Lane Road Comp.	Total Comp.
1	\$12,000	15	\$578	\$7,967	\$20,545
2	\$12,000	16	\$617	\$7,967	\$20,584
3*	\$12,000	13	\$501	-	\$12,501
4	\$12,000	18	\$694	\$7,967	\$20,661
5	\$12,000	17	\$655	\$7,967	\$20,622

"L.F." = Lost Frontage. "Comp." = Compensation. Property 3 Only Property Not on 4 Lane Road.

Model 2: Four Lane Road Model

Compensation = \$12,000 (*Lump Sum*) + \$40.68 × Lot Frontage Lost + \$.84 (33,000 - old traffic count)

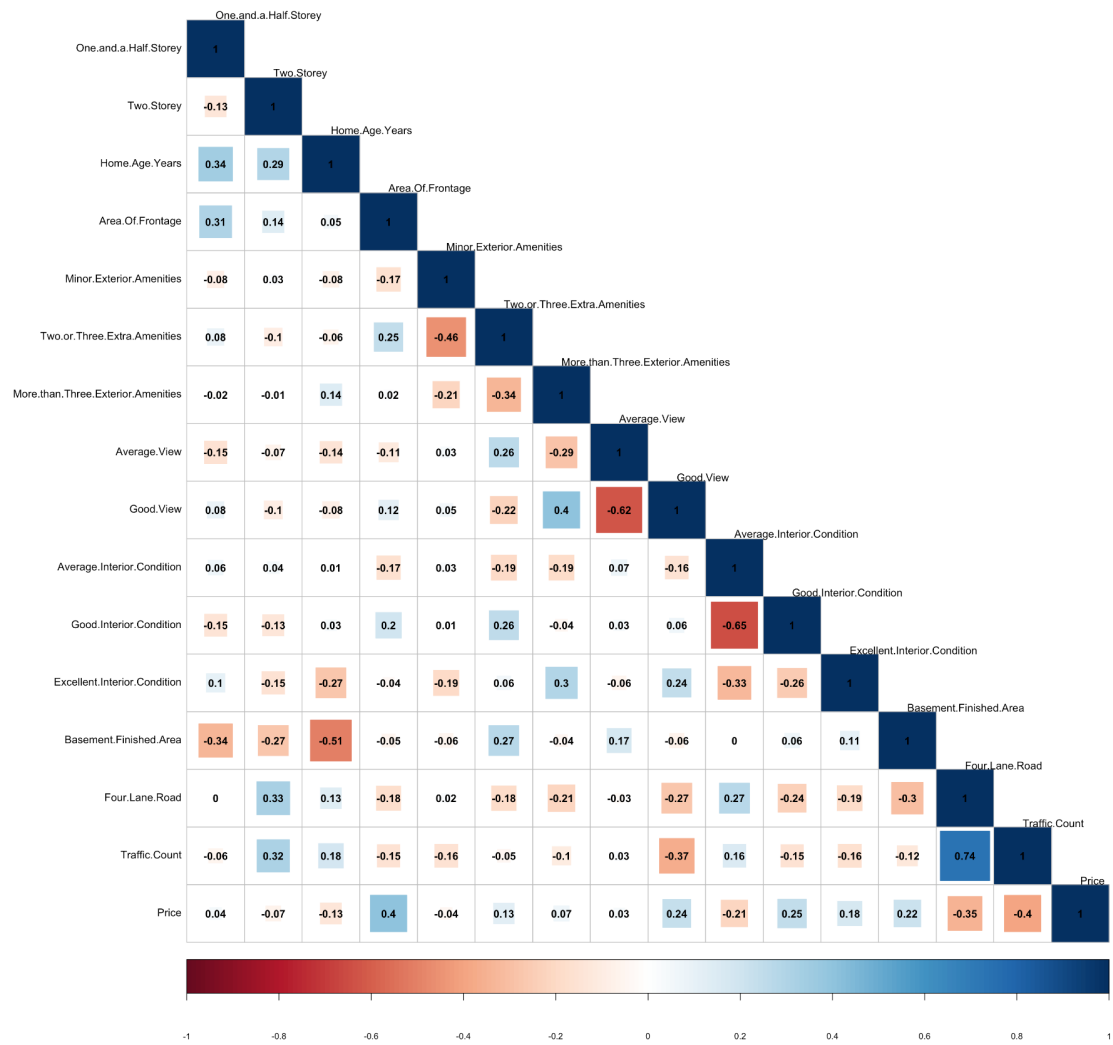
Property #	Lump Sum	L.F.	LF Comp.	Traffic Comp.	Total Comp.
1	\$12,000	15	\$610	\$10,080	\$22,690
2	\$12,000	16	\$651	\$7,560	\$20,211
3	\$12,000	13	\$529	\$5,880	\$18,409
4	\$12,000	18	\$732	\$7,560	\$20,292
5	\$12,000	17	\$691	-	\$12,691

"L.F." = Lost Frontage. "Comp." = Compensation. Property 5 Not Directly Impacted Traffic-Wise.

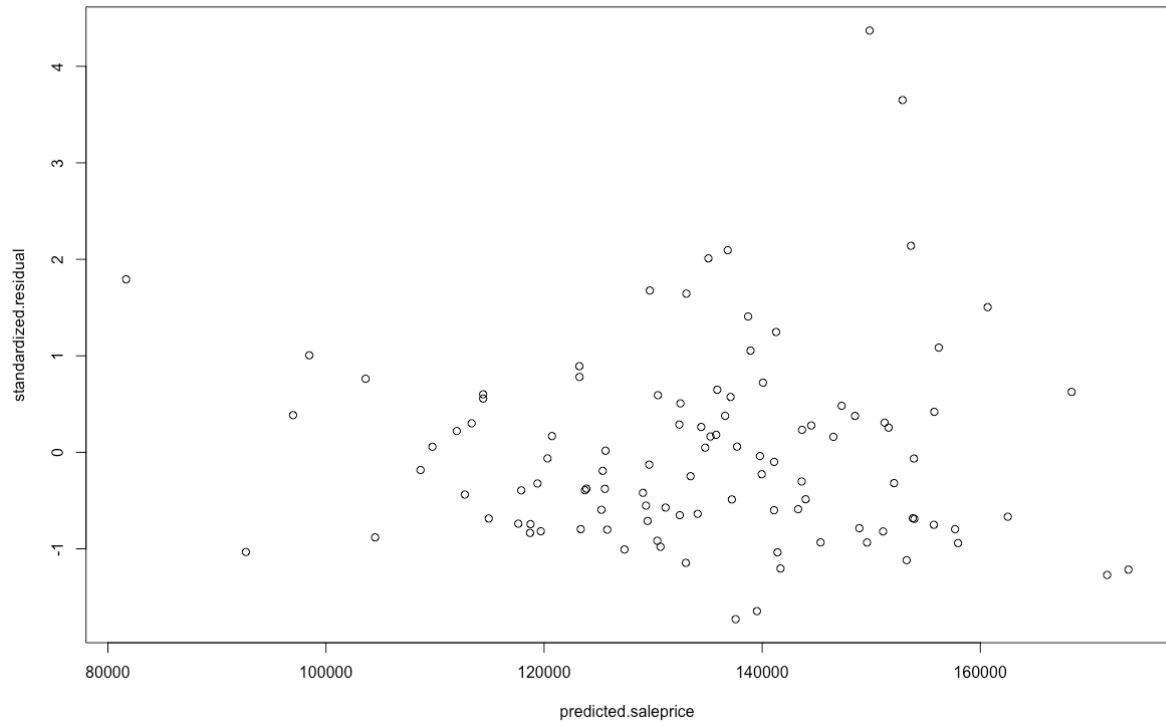
Limitations

One limitation of the model is the sample size relative to the number of potential variables that could be included in a multiple regression model. With a sample size of 104 observations, in addition to the variables in question being area of frontage, and traffic count / number of lanes, only 8 additional total variables can be included if the model is to be considered 'normal'. This is a severe limitation as the number of variables that affect home price are almost certainly larger than 8. In fact, with a multiple R squared of .373 and Adjusted R square of .275, it's clear that the number of variables left on the table are potentially double the number of the 8 included.

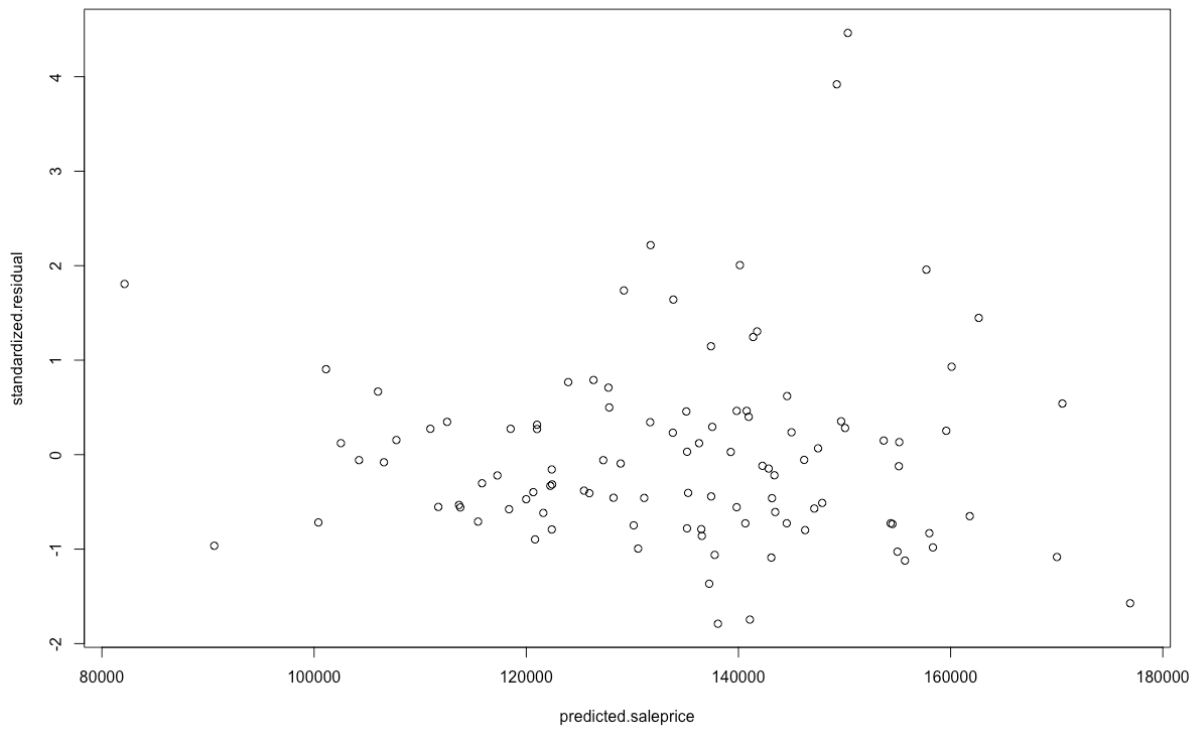
Exploratory correlation plot for the variables.



Four Lane Road Model: Residuals



Traffic Count Model: Residuals



Word Count: 1,499