Land Use Efficiency: How Unit Count and Stories Impact Lot Value*

Ziheng (Tony) Fang

Abstract

For properties on the real estate market, one may expect that the larger the lot size, the larger the market value of the property. In the real world, that may be true, however, would the relationship between lot size and assessed value of the property alse be impacted by the total number of units the property contains and the number of stories the property hosts? Some in the real estate industry would wonder if having more residential units or a higher building would increase or decrease the value of the land the property is built on. This is an extremely important question for many urban environments as there is extreme population density and housing is scarce. Land use efficiency is crucial in these markets. Data regarding total assessed value of different residential properties was studied from real estate data provided by the Assessor's Office in San Francisco. Methods utilized included building a regression model, studying the interaction effects, and running a hypothesis test for the interaction effects.

Introduction

In many urban areas across the country, the population is extremely dense, making land among one of the most scarcest and valuable resources for real estate developers. In the eyes of a developer, how efficiently land is utilized, whether it's having multiple units in one property or multiple stories in a unit, significantly impacts the city's housing supply. In cities like San Francisco, where housing is limited and prices are up, land use is a significant problem that city leaders want to address. On top of supply, understanding how lot area would interact with the number of units inside a property and the number of stories a unit has can provide additional insights on how efficient land use can actually have an impact on property value in the eyes of a potential homeowner. From this our hypothesis test in this study, would also be able to see whether or not there is a positive correlation between lot area and the assessed value.

*Project GitHub Repository: https://github.com/TFang96/MATH261A-Project2

Prior research often examined how the prices of property are predicted by direct predictors such as the lot area, room count, bedroom count, etc. However, these studies very rarely address how variables regarding efficient land use, such has having multiple units in a building and multiple stories in a unit could impact how our predictors affect property value. In reality, many of these variables are interdependent indicating that while we could utilize land more efficiently by having a residential unit with multiple floors and having multiple residential units in one building, in the eyes of a property owner, the value per unit of lot size may also change significantly.

This paper uses data from the San Francisco Assessor-Recorder office to see how property values are impacted not only by the lot area, but by how the number of units inside a property and the number of stories per unit impacts how lot area affects property values. A multiple linear regression model was utilized to examine both the direct effect of lot area on the assessed value and how the unit count and story count affects the impact lot area has on the assessed value through the interaction effect.

By identifying how vertical expansion and having multiple units inside a building affects how a unit change in lot area affects property value, this paper provides additional insight for developers when considering land use. Insights from this analysis can help developers and policymakers better understand trade-offs when it comes to land use efficiency and real estate value.

The Data section gives further details on data that was utilized for this paper. Details on the statistical methods utilized are listed in the Methods section. The Results section includes our statistical findings. The conclusion we drew can be found in the Discussion section.

Data

The data set used for this paper, Assessor Historical Secured Property Tax Rolls (*Assessor Historical Secured Property Tax Rolls* 2025), provides details such as the assessed values, lot area, number of units, number of stories, and more for properties within the city of San Francisco since 2016. The data was compiled by the city assessor's office. We have a total of 3.72M observations in the original dataset where each observation represents a real estate property. We filtered the data to only include observations from one year, 2024, to prevent an issue where we counted a single property more than once. Variables that we used for the response variable in our study included Assessed.Fixtures.Value, Assessed.Improvement.Value, Assessed.Land.Value, and Assessed.Personal.Property.Value, which were summed up into the AssessedValue. Variables used as our predictor variables included Lot.Area, Number.of.Units, and Number.of.Stores. To allow numerical computations, all commas were removed. Since clients look for very different features when on the market for residential and commercial properties, we only kept data on single family residential and multiple family residential units.

Table 1: Summary statistics for assessed property values, unit count within each building and story count within a building.

	Value
Mean_Assessed_Value	1.255264e+06
Assessed_Value_SD	4.848550e+06
MinAssessedValue	0.000000e+00
MaxAssessedValue	4.616426e+08
Mean_LotArea	3.833490e+03
LotArea_SD	1.588195e+05
Min_LotArea	1.200000e-01
Max_LotArea	5.800145e+07
Mean_NumberOfUnits	2.250725e+00
NumberOfUnits_SD	9.463662e+00
Min_NumberOfUnits	0.000000e+00
Max_NumberOfUnits	7.600000e+02
Mean_NumberOfStories	1.719294e+00
NumberOfStories_SD	6.418415e+00
Min_NumberOfStories	0.000000e+00
Max_NumberOfStores	9.720000e+02

Methods

We started off by fitting a linear regression model based on the dataset using the lm() function in R (R Core Team 2024). The original model is as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} X_{i1} + \beta_3 X_{i1} X_{i2} + \epsilon_i$$

In the original regression model, our predictor variables include the lot area of a property (X_{i1}), the number of units (X_{i2}), and the number of stories (X_{i3}). The response variable (Y) is the total assessed value of that property. The intercept (β_0) represents the hypothetical assessed value of a property that has a lot area of 0 sq. ft, 0 units, and 0 stories. The coefficient for lot area (β_1) represents the change in assessed value per unit increase in lot area. The coefficient β_2 represents the interaction effect between lot area and the number of units, or the change in the effect of lot area for every extra unit in a property. The coefficient β_3 represents the interaction effect between lot area and the number of stories, or the change in the effect of lot area for every extra story in a property. We also have an error term, ϵ_i which includes variations in assessed value that are not influenced by our predictor variables.

However, upon examining the residual vs. fitted values plot, we can observe extreme heteroskedasticity, reflecting unequal variance among the error terms, which makes our hypothesis tests off. To address this, we applied a log transformation to the response variable, where the transformed model shows substantially more constant variance.

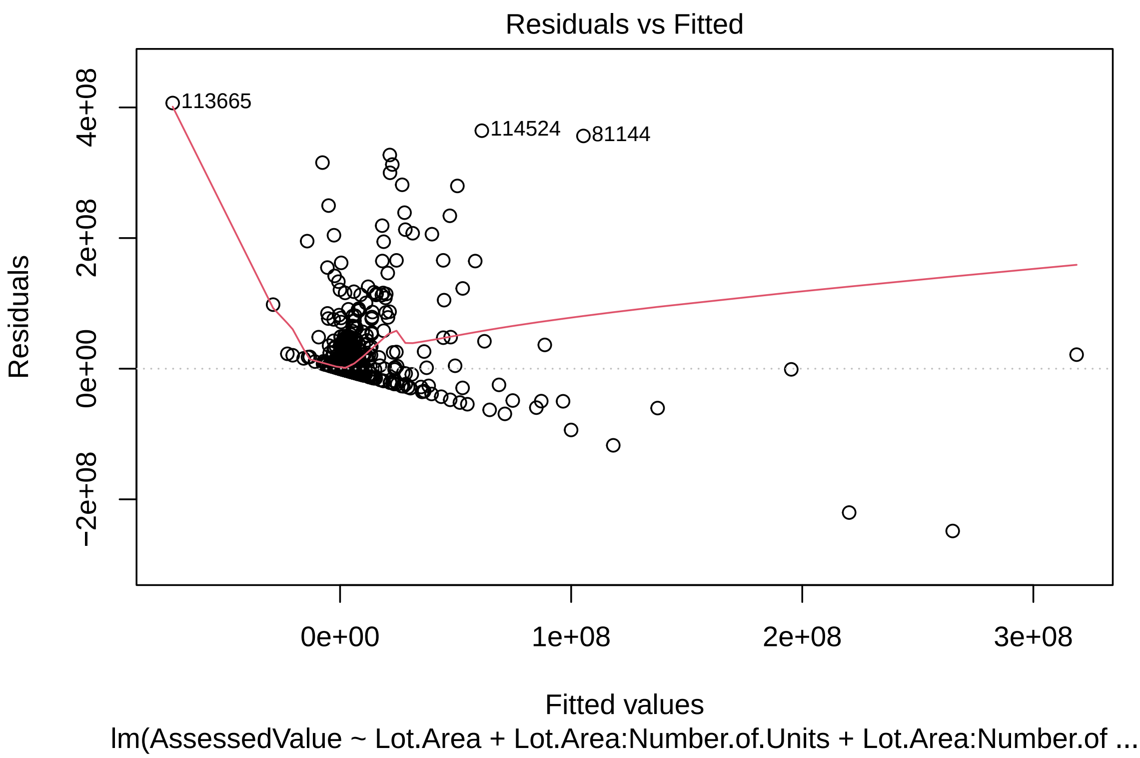

Figure 1: Residual vs. Fitted Values in original model. Observe extreme heteroskedasticity.

The transformed model is as follows:

$$log(Y_i + 1) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} X_{i1} + \beta_3 X_{i3} X_{i1} + \epsilon_i$$

In our transformed model, our predictor and response variables remain the same. However, when it comes to our coefficients, the intercept (β_0) represents the predicted log-assessed value when all other predictors are equal to 0, which is not realistic in our context, the coefficient for lot area (β_1) represents the percentage change in the assessed value for a unit increase in lot area when all other predictors are 0, the coefficient for the interaction effect between lot area and the number of units (β_2) represents the percentage change in the percentage change in assessed value per unit increase in lot area for every extra unit in the property, and the coefficient for the interaction effect between lot area and the number of stories (β_3) represents the percentage change in the percentage change in assessed value per unit increase in lot area for every extra story in the property.

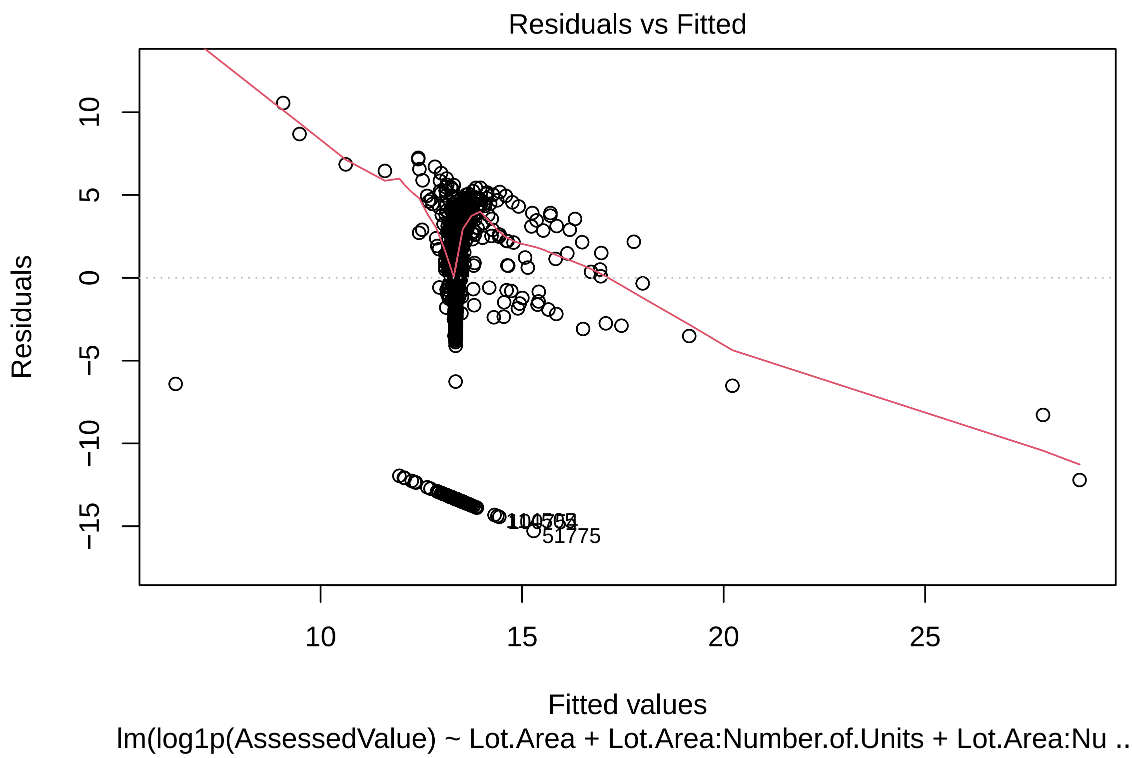

Figure 2: Residuals vs. fitted values in log transformed model. Observe reduced heteroskedasticity.

As indicated by figure 2, when we compare the amount of heteroskedasticity in both models, it is significantly reduced in the log model.

Results

For each coefficient (β_1, β_2, β_3) in the log model, we ran a t-test as our hypothesis test with a test size of 0.05.

For β_1, which represents the percentage change in assessed value per unit increase in lot area, we had a null hypothesis H_0, which assumed β_1 to be equal to 0. This would indicate that there is no relationship between lot area and the assessed value. Our alternative hypothesis,

H_a, would assume that β_1 was not equal to 0, indicating a relationship between lot area and assessed value. The test statistic we used was:

$$t = \frac{b_1 - 0}{SE(b_1)}$$

We observed a p-value of less than $2 * 10^{-16}$, which leads us to reject the null hypothesis as it is less than our test size. Observing an estimated coefficient of $-4.032 * 10^{-6}$, we can estimate that an extra square foot in lot size leads to a decrease in assessed value by $4.032 * 10^{-4}\%$.

For β_2, which represents the change in the percentage change in assessed value per extra square foot of lot area for each extra unit in the building, we also had a null hypothesis H_0, which assumed β_2 to be equal to 0. This would indicate that the rate of change in the assessed value per extra square foot in lot area is not affected by extra units in the property. In contrast, the alternative hypothesis H_a, would assume that β_2 is not equal to 0, indicating that the rate of change in the assessed value per extra square foot in lot area is affected by extra units in the property. The test statistic we used was:

$$t = \frac{b_2 - 0}{SE(b_2)}$$

We observed a p-value of less than 0.000657, which leads us to reject the null hypothesis as it is our test size. Observing an estimated coefficient of $-5.323 * 10^{-9}$, we can estimate that each extra unit inside a property leads to the percentage change in the assessed value per extra square foot in lot area to decrease by $5.323 * 10^{-7}\%$.

For β_3, which represents the change in the percentage change in assessed value per extra square foot of lot area for each extra story in the building, we also had a null hypothesis H_0, which assumed β_3 to be equal to 0. This would indicate that the rate of change in the assessed value per extra square foot in lot area is not affected by extra stories in the building. Whereas, our alternative hypothesis H_a, would assume that β_3 is not equal to 0, indicating that the rate of change in assessed value per unit increase in lot area is impacted by an additional story. The test statistic we used was:

$$t = \frac{b_2 - 0}{SE(b_2)}$$

We observed a p-value of less than $2 * 10^{-16}$, which again leads us to reject the null hypothesis. Observing an estimated coefficient of $2.018 * 10^{-6}$, we can estimate that each extra story on a property leads to the percentage change in the assessed value per extra square foot in lot area to increase by $2.018 * 10^{-4}\%$.

Discussion

To look into how common methods used to efficiently use land for housing can impact how an additional square foot of lot area affects the assessed value of housing properties in San Francisco, we analyzed 133816 residential units. The total assessed value was considered to be the response variable of lot area, as the predictor variable, with units inside the property and the story count of the property as interaction effects.

To determine how additional units in square feet of lot area impacted the assessed value, and how additional units inside the property and additional stories affected the impact lot area had on the assessed value, we ran a t-test as our hypothesis test with a test size of 0.05. For each respective coefficient, we considered a null hypothesis which would indicate that there is no relation between lot area and the assessed value or that a particular variable had no impact on how an additional unit of lot area affected the assessed value. The alternative hypothesis would conclude that each additional unit of lot area did have an impact on the assessed value or that a particular variable did impact how a unit of lot area affected the assessed value through an interaction affect. The estimated coefficients would then tell us about how much of an affect an additional unit of lot area had on the assessed value or how much a variable affected the impact of an extra unit of lot area on the assessed value through an interaction affect.

When we looked into how an additional square foot of lot area affects the assessed value, we observed a p-value that is significantly smaller than our test size. Thus, we rejected the null hypothesis and concluded that the lot area is related to the assessed value of a property. The estimated coefficient also tells us that considering no units inside a property and no stories, we can expect the assessed value to decrease by 0.0004032% for every square foot of lot area. Contrary to popular belief, just increasing the lot area would actually decrease the assessed value.

When we consider how additional units inside a building would impact how each additional square foot in lot area affects the assessed value, we observe a p-value that is also significantly smaller than our test size. In this case, we also reject the null hypothesis and conclude that the number of units a property has will impact how each additional square feet in lot area affects the assessed value. The estimated coefficient tells us that considering no stories, every extra unit inside a building will decrease the change in assessed value per additional square foot in lot area by 0.0000005323%.

If we were to consider how additional stories to a building would impact how each additional square foot in lot area affects the assessed value, by looking at the p-value, we also observe a value that is significantly smaller than our test size. From that we would reject the null hypothesis and conclude that the number of stories a property has will impact how each additional square feet in lot area affects the assessed value. The estimated coefficient tells us that considering no units, every extra story on a building will increase the change in assessed value per additional square foot in lot area by 0.0002018%.

For a developer looking into how they can efficiently utilize land while also maximizing profit in real estate, this study would suggest that by building taller buildings will not only be an efficient use of land, but will have a positive effect on how each additional square foot of lot area used affects the assessed value. On the other hand, while having multiple units inside a building can save land area, it would have a negative effect on how each additional square foot of lot area used affects the assessed value.

However, this research has a limitation that the dataset used only contains properties in San Francisco, an urban city with significant population density. In an urban area, it can be reasonable to see that a taller building might positively affect how an additional square foot of lot area impacts the assessed value, simply because we are trying to save space. However, in more rural areas, many homeowners might prefer their property to be on one story. To make our findings more conclusive for properties regardless of location, we could take samples of real estate properties in urban, suburban, and rural areas.

References

Assessor Historical Secured Property Tax Rolls. 2025. Assessor's Office, City of San Francisco. https://data.sfgov.org/Housing-and-Buildings/Assessor-Historical-Secured-Property-Tax-Rolls/wv5m-vpq2.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.