Estimating the Market Value of House Remodels

Lab 2 Sample Answer - Paul Laskowski

Introduction

As today's housing stock ages, the choice to remodel a property is driven by many factors, some personal and some market-wide. For many owners, a major concern is the effect a remodel will have on future revenues, including those from renting and from selling the property. Some may turn to real estate professionals for guidance. In a 2019 survey performed by the National Association of Realtors, members estimated the financial benefit of different type of remodels, ranging from $2,500 for a closet renovation up to $20,000 for a full kitchen remodel.

While experts may provide broad guidelines, data-based approaches are needed to reduce uncertainty in the value of remodels. In the aggregate, remodeling accounts for $400 billion spent each year in the US.[1] Uncertainty in how much of this money can be recouped may contribute to the misallocation or underprovision of resources in this sector of the economy.

This study estimates the economic value for remodeling a home empirically, utilizing observations of house sales in Ames, Iowa. The data shows the timing of the last remodel before a house is sold, but does not distinguish between different types of remodels. Applying a set of regression models, I estimate the value that results immediately when a house is remodeled, and also the rate at which it decays over time.

Data and Methodology

The data in this study comes from the Ames, Iowa Assessor. It was compiled and made publicly available by Dean De Cock.[2] Each row in the data represents a home sale in Ames between 2006 to 2010. I performed all exploration and model building on a 30% subsample of the data. The remaining 70%, totaling 2051 rows, was used to generate the statistics in this report.

Detecting which homes are remodeled requires some care. The year built variable represents the original construction date, while the year remodeled variable contains the time of the last remodel before the sale. According to the data dictionary, if a home was never remodeled, the year remodeled and year built should be the same. Because the year remodel variable is censored at 1950, there is no way to reliably determine which homes built in 1950 or earlier were never remodeled. I filter out the 477 homes in this category, leaving 1574 observations. There are also 282 homes recorded as remodeled one year after being built. In fact, this includes 0.35 of homes built after 2000. Because that fraction is implausibly high, I interpret these records as showing two dates for what is really a single construction. I therefore consider a house to be remodeled if there is at least a 2-year difference between the year build year and the year remodeled. As a robustness check, I reran the analysis using a 1-year difference, and found that this does not cause any substantial change in the key regression coefficients.

To operationalize economic value, I divide the sale price of a home by the number of square feet of above-ground living area. This form was chosen to best fit the relationship seen in exploratory plots. Figure 1 plots price per square feet as a function of age, with remodeled homes shown in blue and non-remodeled in red. There is a general negative relationship between value and age. Interestingly, among homes that are less than 15 years old, remodeled homes are worth less than non-remodeled homes on the average.

[1] La Jeunesse, Elizabeth. "Healthy Home Remodeling: Consumer Trends and Contractor Preparedness." Harvard's Joint Center for Housing Studies (2019).

[2] De Cock, Dean. "Ames, Iowa: Alternative to the Boston housing data as an end of semester regression project." Journal of Statistics Education 19.3 (2011).

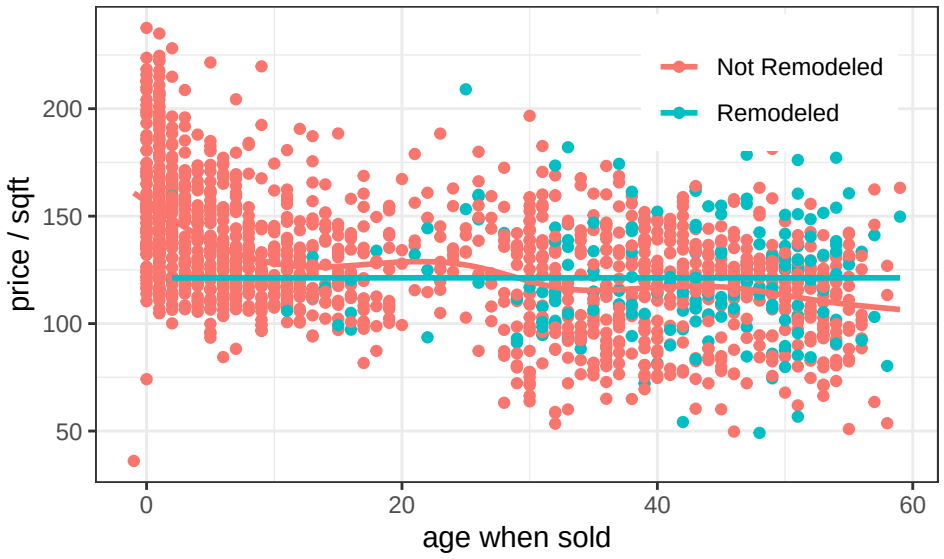

Figure 1: Home Value as a Function of Age

I am interested in the difference in value between two counterfactuals: one in which a house is remodeled, and another in which it is not. There are two key factors that might affect this increase in value:

- The age of the house when remodeled. It is possible that remodeling an older home results in a larger value increase than remodeling a newer home.
- The time between the remodel and the sale. It is possible that remodels become more or less valuable over time, and the rate of change may be different than that of houses in general.

Exploratory plots suggest that both effects exist and that both are roughly linear. I therefore create regression models in which the "boost" from remodeling increases by a fixed amount with the age of the house when remodeled, and also changes by another fixed amount with each year that passes after the remodel. In other words, I fit regressions of the form,

$$\widehat{price/sqrtft} = \beta_0 + \beta_1 \cdot R \cdot (year\ remodeled - year\ built) + \beta_2 \cdot R \cdot (year\ sold - year\ remodeled) + \mathbf{Z}\gamma$$

where R is an indicator for remodeling, β_1 represents the immediate increase in value per year the house existed before remodeling, β_2 represents the change in the value increase for each year that passes after the remodel, $\mathbf{Z}$ is a row vector of additional covariates, and γ is a column vector of coefficients.

I considered specifications that also include the modeling indicator R by itself (i.e. uninteracted). This type of model allows for the possibility that even a brand new house that is remodeled immediately increases in value. However, when fitting such models in the exploration set, the resulting coefficient was practically small (equivalent to reducing the age of a home by 1 to 2 years) and non-significant. To improve the precision of my estimates and the simplicity of the model, I removed this term.

Results

Table 1 shows the results of three representative regressions. Across all models, the key coefficient on $R \cdot (year\ remodeled - year\ built)$ was highly statistically significant. Point estimates range from 0.26 to 0.42. To provide some sense of scale, consider a hypothetical house with 2,000 square feet of living area that was never remodeled before. Applying model 3, if the house is 20 years old, a remodel is predicted to increase resale value by 10,000 dollars. If the house is 50 years old, the number rises to 26,000 dollars.

Across all models, the coefficient on $R \cdot (year\ sold - year\ remodeled)$ was also highly statistically significant. Point estimates range from -0.48 to -0.35. This is evidence that the boost in value from a remodel fades over time. Moreover, the extra value decreases even relative to the overall effect of aging for a house. For an

Table 1: Estimated Regressions

	Output Variable: price per square foot		
	(1)	(2)	(3)
$R \cdot (year\ remodeled - year\ built)$	0.42^{***}	0.28^{***}	0.26^{***}
	(0.07)	(0.05)	(0.05)
$R \cdot (year\ sold - year\ remodeled)$	-0.48^{**}	-0.36^{***}	-0.35^{***}
	(0.16)	(0.11)	(0.10)
age	-0.71^{***}	-0.85^{***}	-0.72^{***}
	(0.04)	(0.07)	(0.08)
lot square feet		0.0004^{***}	0.0005^{***}
		(0.0001)	(0.0001)
living square feet		-0.01^{***}	-0.02^{***}
		(0.003)	(0.003)
full baths		-9.27^{***}	-7.98^{***}
		(1.82)	(1.81)
half baths		-11.26^{***}	-10.10^{***}
		(1.35)	(1.29)
total rooms		-3.58^{***}	-3.10^{***}
		(0.68)	(0.64)
year sold		-3.83	-4.32^{*}
		(2.20)	(2.14)
constant	144.55^{***}	7,842.12	$8{,}798.56^{*}$
	(1.12)	(4,419.08)	(4,306.08)
zoning type		✓	✓
building type		✓	✓
sale type		✓	✓
neighborhood market trends		✓	✓
additional features			✓
Observations	1,574	1,574	1,574
R^2	0.20	0.64	0.69
Residual Std. Error	26.28 (df = 1570)	18.14 (df = 1495)	16.89 (df = 1453)

Note: *p<0.05; **p<0.01; ***p<0.001
HC_1 robust standard errors in parentheses.
Additional features are deck area, open porch area, enclosed porch area, 3-season porch area, street type, lot shape, lot configuration, land slope, masonry veneer type, heating type, central air conditioning, electrical system, roof style, roof material, kitchen number, paved driveway, and pool area.

example, again consider a hypothetical house with 2,000 square feet of living area. According to model 3, if a remodel occurred 5 years ago, it is worth 3,500 dollars less than a remodel today. If the house was 20 years old when remodeled, at 15 years after remodeling, the value boost is predicted to disappear, meaning that the house is worth the same as an equivalent house that was never remodeled. A perhaps unrealistic feature of the model is that the predicted value of a remodel is predicted to become negative after this point.

Limitations

Consistent regression estimates require an assumption of independent and identically distributed (iid) observations. Because homes exist in a geography, there is a possibility of geographical clustering. I partly account for this possibility in models 2 and 3, by including a fixed effect for each neighborhood that is interacted with year sold. In other words, each neighborhood has a unique slope and linear trend over time. I am not able to account for geographical clustering within each neighborhood.

Because house sales take place over time, there is a further possibility of temporal autocorrelation. Real estate professionals often use past sale prices to help value current properties, so a high sale price at one time may increase the probability of high sale prices at future dates.

Consistent regression estimates also require that the population distribution is described by a unique best linear predictor. Supporting this assumption, I do not see any visual evidence of heavy tailed distributions in any diagnostic plot. Variables are automatically dropped to avoid perfect collinearity.

As far as structural limitations, several omitted variables may bias my estimates. In a classic omitted variables framework, the omitted variable is assumed not to interact with the key variable in the true model. An example of a variable for which this assumption is plausible is uneven floors. If uneven floors make it more difficult to perform a remodel, I expect a negative correlation between uneven floors and my key variables. Since uneven floors are likely to have a negative effect on price in the true model, I predict a positive omitted variable bias on the key variables. The main effect is therefore being driven away from zero, making my hypothesis tests overconfident. A similar analysis holds for unusual room dimensions and exposure to earthquakes.

The standard textbook analysis must be modified when a remodel is performed to correct a problem with a house. For an example, consider homes containing toxic materials. While the presence of toxic materials may cause a remodel, it is also an outcome variable. That is, remodeling may cause toxic materials to decrease. In the extreme case, all homes with toxic materials become remodeled, such that no toxic materials remain in the data. The positive benefits of removing toxic materials then become impossible to measure. As a result, I expect a remodel to appear less valuable than it really is. The main effect is therefore being driven towards zero, suggesting that my hypothesis tests are underconfident. A similar analysis holds for problems like unfashionable cabinets, mold, and pet damage.

Conclusion

This study estimated the economic value of remodeling a home. For every year that a house existed before a remodel, our models predict an increase in price per square foot between 0.26 and 0.42 dollars. We also describe and measure an effect in which the value from a remodel tends to disappear over time. The time scale over which the benefit disappears is roughly comparable to the age of the house when remodeled.

In future research, new datasets may be generated to estimate the value of specific types of remodels. Owners may want to know, for example, the financial benefit of new floors or hidding kitchen appliances. The ultimate hope of this line of work is to provide accurate tools for owners to plan investments, and reduce uncertainty in the remodeling industry.