

# Estimating the Market Value of HOA Prices

Jessica Stockham, Derrick Chan-Sew, Jammy Chan (Chi Hung)

## Introduction

Real estate prices have seen a recent surge in pricing in an increasingly competitive market. While there is transparency around overall market value, many homeowners are concerned about the factors that will have a direct effect on the value of their homes. These factors are pivotal in the home buying process and ultimately create long term value for homeowners. Many communities have homeowners associations (HOAs) which present increasing annual costs for homeowners due to inflation. While the homeowners associations are essential in managing common property and providing services to residents, one of their principal functions is to protect property values. Our research focus is how the price of HOAs affect overall property value.

Government officials who derive property taxes based on a home's assessed value, HOA community leaders and the homeowners have equal interest in an empirical analysis in the effect of HOA prices and home value. Homeowners in the community, most likely a subgroup such as the HOA Committee or Board of Directors can define the rate of increase of HOA fees and potential affect home prices. HOA takes care of all common areas, using dues collected from residents to repair, maintain, or replace community amenities and it does affect property value over time.<sup>1</sup> This research creates and evaluates a set of regression models then uses the selected model to estimate home value from HOA paid for a property.

## Data and Methodology

The data in this study comes from the US Census 2021 American Housing Survey (AHS), a survey of U.S. households. We apply the survey weights to our regression models so that the data represents a nationally-representative sample. The observations are at the household-level, including the market value of the home, monthly HOA amount, housing characteristics, neighborhood information, and household socioeconomics. While the full sample includes **64,141** observations, our analytic sample is homeowners (rather than renters) who pay HOA fees and who have a housing unit that is a single-family detached structure, a single-family attached structure, or is part of an apartment building. This excludes mobile homes, trailers, boats, RVs/vans. Based on these filters, our initial analytic sample has **8,010** observations. We also consider a variety of other variables to include as controls in the regression models, all of which have a small amount of missing values. Once we remove observations that are missing values on the controls, our analytic sample is **7,907**.

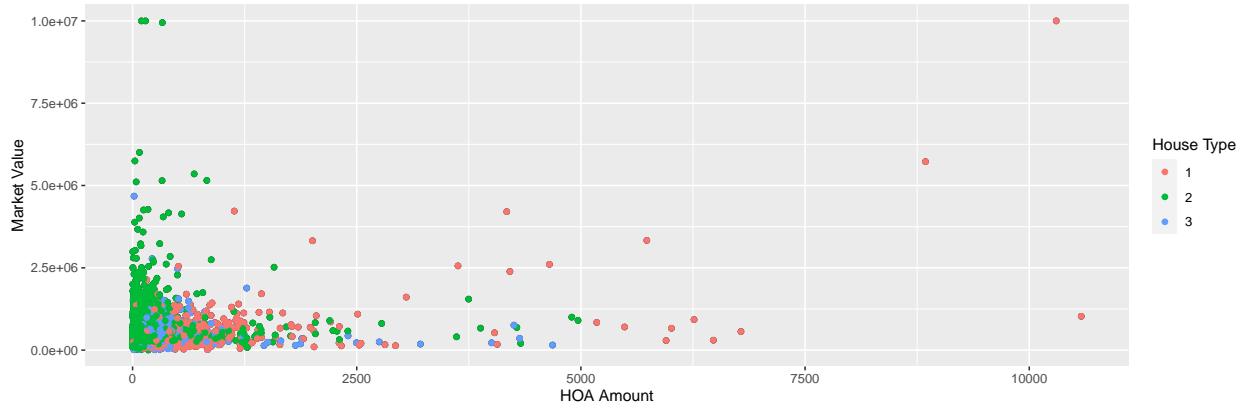
We split the analytic sample into two pieces. We performed all exploration and model building on a 30% subsample of the data. The remaining 70%, totaling **5580** rows, was used to generate the regression results in this report.

Our outcome variable is the owner's report of the market value of the home in dollars and our key independent variable is the monthly HOA fee in dollars. Both are ratio, metric variables that match our desired concepts well. However, they are self-reported, which introduces some error into the data, as further described in the limitations section. Histograms show that both market value and HOA fees are skewed distributions (skewness of **8.26** and **9.98**, respectively), best modeled with log transformations. The histograms are in the appendix.

---

<sup>1</sup>Cedar Management Group, "How Does An HOA Increase Property Values?", <https://www.cedarmanagementgroup.com/hoa-increase-property-values/>, 2022

We believe that the HOA fee may impact market value differently, depending on the type of housing structure you live in. For example, the HOA fee for an apartment building may pay for costly shared building expenses like a new roof and/or a regular maintenance and cleaning service for common spaces. On the other hand an HOA fee for a single-family detached housing complex may have substantially less burden, perhaps only funding some common fencing around the exterior of the neighborhood or lighting at the entrance of the neighborhood. A single-family attached complex, such as a duplex may be somewhere in the middle. As such, all our regression models estimate the interaction between HOA and housing type (apartment, single-family attached home, single-family detached home). We generated a scatter plot to observe the relationship between HOA Amount and Market Value. The colors represent the three different house types. It is difficult to discern a clear relationship, however apartments have more of a linear relationship with market value than the other house types.



In addition to housing type and HOA fees, we believe several other key factors predict the current market value of the home. It is important to include any covariates that are correlated with both HOA fees and the outcome variable to minimize omitted variable bias. As such, we included following candidate covariates in our analysis: (1) basic home characteristics, including year built, square footage, the presence of a garage, and lot size. We also use the owner's rating of the unit, which should account for hard-to-measure factors for which we do not have data. (2) neighborhood features, including the owner's rating of the neighborhood, if there is petty crime, serious crime, quality of the schools, flood zone risk and (3) supply and demand forces in a local geographic area. We expect geographic clustering, for example higher prices of homes in urban centers with high demand. Notably, we ultimately left several of these variables off our regression models due to high multicollinearity as follows: total number of rooms, number of bedrooms, number of bathrooms, owner's rating of the unit.

We built four regression models. First, we used a base model that measures the predictive power of HOA fees (interacted with home type) on the market value of the home. Then we estimated models that incorporated the following groups of variables in turn: basic home characteristics, neighborhood features, and geography, as described above. There is more detail on the definition of these covariates in our appendix. We constructed all covariates as categorical variables/dummies, so that we did not impose a linear relationship with the market value.

## Results

Table 1 shows the results of four models, with robust standard errors applied. Across all models, the  $\ln(\text{HOA Amount})$  remains positive and statistically significant. The point estimates are between 0.18 and 0.28, meaning that if we increase the HOA Amount by 1%, we expect the market value to increase by about 0.2%, all else held equal

Importantly, the interaction effects of  $\ln(\text{HOA Amount}) * (\text{Detached single family homes})$  and  $\ln(\text{HOA Amount}) * (\text{Attached single family homes})$ , compared to apartment buildings, are also statistically significant and negative across all models. In fact, with point values between -0.15 and -0.10, the main effect of HOA Amounts is dampened or largely wiped out for single family homes.

Table 1: Estimated Regressions

	Output Variable: Market Value			
	(1)	(2)	(3)	(4)
HOA Amount	0.28*** (0.06)	0.24*** (0.05)	0.24*** (0.05)	0.18*** (0.05)
Single-Family (D)	1.50*** (0.36)	0.79** (0.28)	0.80** (0.28)	0.64* (0.27)
Single-Family (A)	1.20** (0.39)	0.74* (0.30)	0.76* (0.30)	0.66* (0.29)
Garage - No		-0.23*** (0.04)	-0.22*** (0.04)	-0.18*** (0.04)
Near petty crime - No			-0.04 (0.03)	-0.04 (0.03)
Near serious crime - No			0.002 (0.10)	0.07 (0.09)
Near a disaster area - No			-0.13** (0.04)	-0.16*** (0.04)
$HOAAmount \cdot SingleFamily(D)$	-0.15* (0.06)	-0.13** (0.05)	-0.13** (0.05)	-0.10* (0.05)
$HOAAmount \cdot SingleFamily(A)$	-0.19** (0.07)	-0.14** (0.05)	-0.14** (0.05)	-0.12* (0.05)
Constant	11.00*** (0.36)	12.00*** (0.32)	11.00*** (0.39)	11.00*** (0.35)
Lot Size		✓	✓	✓
Square Footage		✓	✓	✓
Year Built		✓	✓	✓
Neighborhood Rating			✓	✓
School Rating			✓	✓
Metro Code				✓
Observations	5,580	5,580	5,580	5,580
R <sup>2</sup>	0.12	0.33	0.34	0.48
Residual Std. Error	27.00 (df = 5574)	24.00 (df = 5548)	24.00 (df = 5534)	21.00 (df = 5518)

Note:

$HC_1$  robust standard errors in parentheses.  
HOA Amount is logged. D = Detached, A  
= Attached.

Practically speaking, if the average HOA Amount is about \$500 a month for an apartment owner who pays HOA fees and the average apartment market price is \$450,000 (as suggested by unweighted AHS data), this means that raising the monthly HOA by \$5 is expected to raise their market price by about \$9,000 ( $\$450,000 * 0.02$ ).

## Limitations

Our OLS regression model is consistent if we (1) assume observations are independent and identically distributed (iid) and (2) if the population distribution is described by a unique best linear predictor (BLP). Housing prices are certainly spatially dependent, violating the independence assumption. We expect that housing prices are clustered geographically, for example certain neighborhoods are more expensive than others. To attempt to address this problem, we apply a set of geographic fixed effects at the regional level, as well as variables to reflect neighborhood characteristics, such as neighborhood rating, school quality, and crime rate. It is likely that we do not fully account for neighborhood clustering with this approach. With regard to the BLP, we do not see any visual evidence of heavy-tailed distributions in any diagnostic plot after we used the log transformation on our metric variables (see Appendix). In addition, our regression models drop any perfectly collinear features. Appendix C also shows our analysis of Classic Linear Model Assumptions on our last model 4, which we show for completeness. While our model fails these tests, our dataset is large enough that we are not constrained by these assumptions.

Since we are limited to the variables in our survey dataset, our estimate of the coefficient on HOA fees may also suffer from omitted variable bias. This bias exists when an omitted variable is correlated both with HOA fees and the house's market value. For example, homes with high HOA fees may also be homes with fancy amenities (e.g., nice gym, rooftop deck, tennis courts). These fancy amenities may raise the market value of the home, but we do not have a robust way to account for them in our survey data. I expect a positive correlation between fancy amenities and HOA fees and a positive correlation between fancy amenities and our outcome variable of market value. Therefore, I expect the main effect is being driven farther away from zero, making my hypothesis overconfident.

In addition, some of the neighborhood variables in our datasets are quite coarse, so they cannot represent the diversity that exists in the country. For example, school quality and crime are binary yes/no questions, such as "Agree or disagree: this neighborhood has good schools." The impact on the model is that we likely have higher error, in other words, we are less predictive of the true population model than we would be with more granular survey questions.

Next, our data is based on self-reports from survey respondents. Self-reports are subject to recall bias, social-desirability bias (e.g., either downplaying the value of the home and the HOA fee or embellishing it), and simply not knowing the true value (e.g., You have the wrong estimate in your head for the square footage of your home).

Finally, while we included survey weights, we did not control for the complex sample design of the survey (requiring replicate weights) in these estimates. As a result, the estimates of the standard errors are likely biased.

## Conclusion

This study assessed the importance of HOA fees in predicting the market value of a home. Due to our study limitations, this is a correlational, rather than a causal study. We found that HOA fees are positively correlated with market value for apartments. There is less evidence of this relationship for single family homes. This is consistent with our theory that HOAs play a more important role in the operation of an apartment than for a single family home. As apartment owners view the next annual increase in HOA payments, they might view them more as an investment in their own property value, rather than a penalty. Alternatively, single-family home owners should be more cautious when approving HOA fee increases. These fees seem less correlated with market value. A future study might look at how changes in HOA fees over

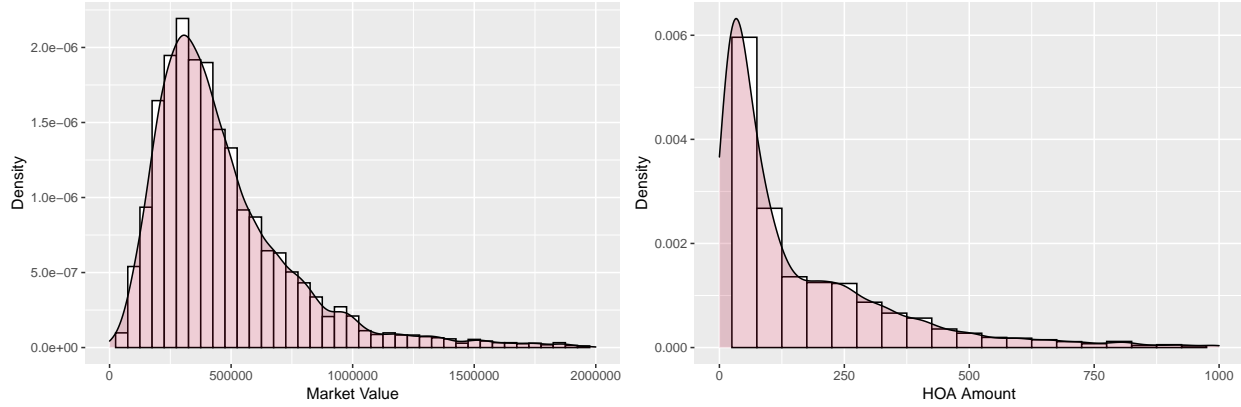
time for the same home affect the market value of the home, rather than a correlational study of a single year of data.

## Appendix A

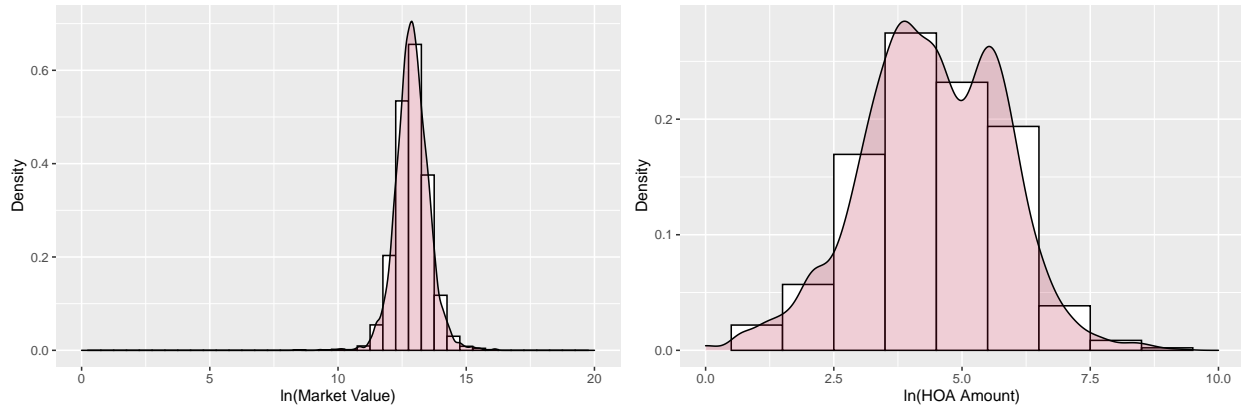
The following variables represent recodes from the raw survey data as follows: House Type (BLD variable recoded to exclude categories 1 and 10. 1 = apartment, 2 = detached single family home, 3 = attached single family home) Year Built: 2011 through 2021 recoded as 2020, Less than 1920 recoded as 1920. Lot Size: No recode. Not that this retains category -6 (Not Applicable) as a valid category. Dropping all -6 would drop all apartments, which we did not want. Garage: Not Reported or Not Applicable coded as NA High Petty Crime: Not Reported or Not Applicable coded as NA High Serious Crime: Not Reported or Not Applicable coded as NA Good Schools: Not Reported or Not Applicable coded as NA Disaster Zone: Not Reported or Not Applicable coded as NA Neighborhood Rating

## Appendix B

### Normality



From the distribution graphs above for both *Market Value* and *HOA Amount*, we found they are skewed to the right. A skewness check confirms this with market value as **8.26** and HOA amount as **9.98**, respectively. To adjust the normality of the data, we decided to apply logarithmic function on both *Market Value* and *HOA Amount*. New transformed variables  $\ln(\text{Market Value})$  and  $\ln(\text{HOA Amount})$  are created respectively.



As shown from the graph above, we found that the graphs shows a more normal distribution compared to the original ones after the transformation. Therefore, we proceed to use the logarithmic variables  $\ln(\text{Market Value})$  and  $\ln(\text{HOA Amount})$  in the linear regression model later.

## Appendix C

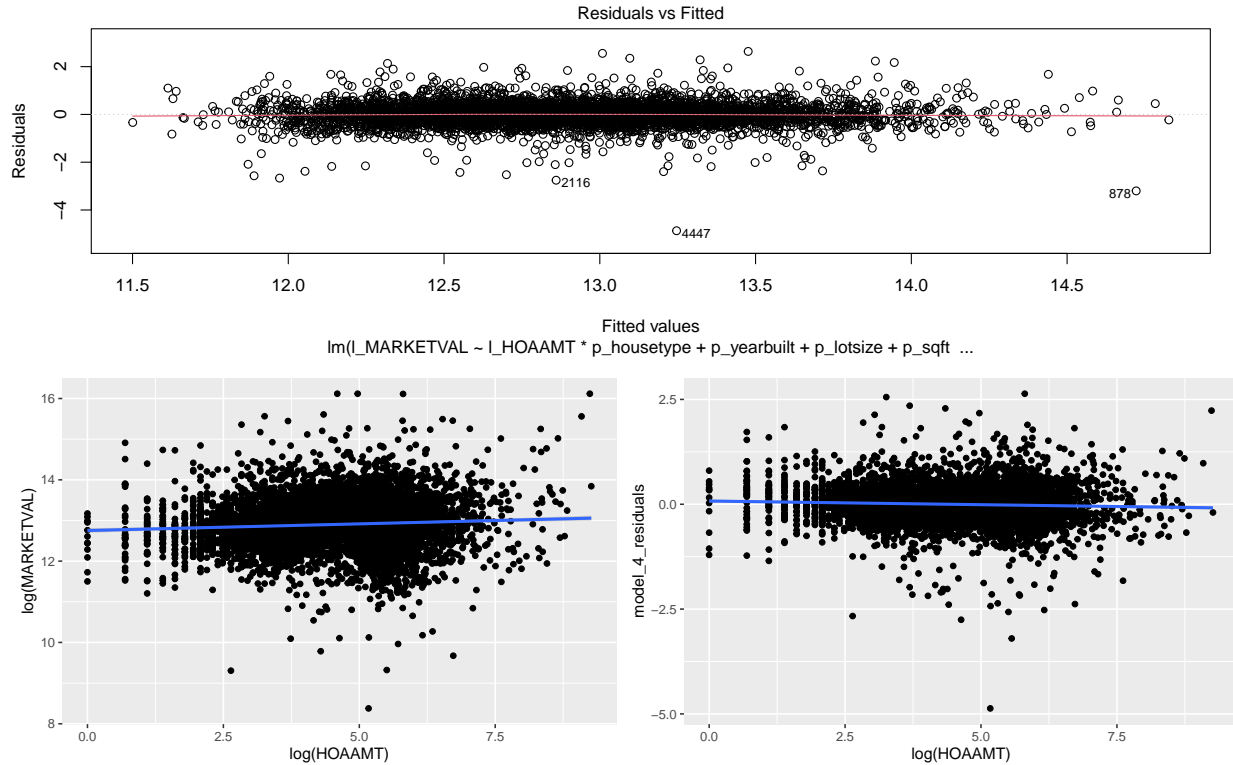
In addition to the big data OLS assumptions of iid and having a unique BLP, we also check three additional Classic Linear Model assumptions: linear conditional mean, homoskedastic errors, and that the errors are normally distributed. These additional assumptions are not necessary for us to have a consistent model, since we are in the big data world, but we provide the results for completeness.

First, we find some evidence that we do not have a *linear conditional mean* with the following three plots: (1) the model 4 residuals versus the predicted values, (2) the log(HOA Amount), our only metric predictor, against the outcome of log(market value) and (2) plotting the log(HOA Amount) against the model 4 residuals. The plots, particular the latter two, appear to have some slight non-linearity to them, but it is not a severe deviation.

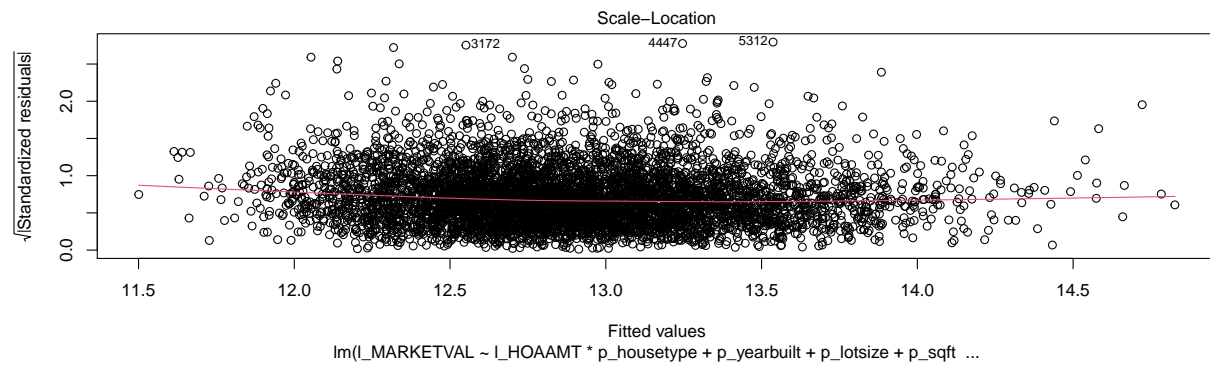
Second, we find evidence that the *errors are not homoskedastic*. We plot standardized residuals vs fitted value and conduct a Bruesh-Pagan test (null is that the data are homoskedastic). The plot hints at some heteroskedasticity with some fanning out and we also reject the null of the Bruesh-Pagan test that the errors are homoskedastic ( $p < 2e-16$ ).

Lastly, we find evidence that the *errors are not normally distributed*, both with a Q-Q plot and a Shapiro-Wilk test (null is that the errors are normally distributed). The plot hints at non-normality and we reject the null of the Shapiro-Wilk test that the errors are homoskedastic ( $p < 2e-16$ ).

### Linear Conditional Expectation: 3 Plots



## Heteroskedasticity Plot



## Normally Distributed Errors Plot

