Analysis Report: Contributions to Current Market Value of Single Family Units in the United States for 2013

Executive Overview

The business question analyzed: What factors predict or contribute to the current market value of single family units? Factors of particular interest in contributing to the current market value of single family housing units for 2013 include the number of rooms in the housing unit, number of bedrooms in the housing unit, location in the western and northeastern region of the US as compared to the southern region, and if the housing unit is in moderately or severely inadequate condition as compared to adequate condition.

Introduction

The business question analyzed: What factors predict or contribute to the current market value of single family units? Multivariate linear regression analysis was conducted to identify any relationship between variables within the Housing Affordability Data System dataset regarding factors that predict or contribute to the current market value of single family housing units. The conclusion of the analysis was that there are many factors that predict or contribute to the housing or current market value of single family units. Factors that have statistical significance and seem to contribute are:

- BEDRMS: number of bedrooms in the housing unit.
- LMED: average median income for the area.
- FMR: cost or fair market monthly rental rate of the housing unit.
- BUILT: the year in which the housing unit was built.
- ROOMS: number of rooms in the housing unit.
- UTILITY: cost of utilities for owning the housing unit.
- REGION: location in the United States of the housing unit.
    - however, there is no statistical significance in market value of the housing unit and the housing unit being located in the Midwest region of the US as compared to the southern region.
- METRO3: location to metropolitan city of the housing unit.
- OTHERCOST: cost or other utility costs for the occupant of the housing unit.
- ZINC2: annual household income of the occupant of the housing unit.
- AGE: Age of head of household of the housing unit.
- ZADEQ: Adequacy of housing unit.
- PER: Number of persons in household in the housing unit.

Body

Question: What factors predict or contribute to the current market value of single family units?

Data

- Data was obtained from the Office of Policy Development and Research (PD&R) U.S. Department of Housing and Urban Development from the Housing Affordability Data System. HADS data "categorizes housing units by affordability and households by income, with respect to the Adjusted Median Income, Fair Market Rent (FMR), and poverty income. It also includes housing cost burden for owner and renter households." ("How is market value determined in the real estate market?", n.d., para. 1) Data was cleaned to include only single-family houses, flats, apartments with Fair Market Value of at Least $1000.00 owned in 2013 as this was the subset that stakeholders wanted to focus on. NULL values represented 2% of the data. All NULL values were removed per stakeholder input. Missing values were missing because the original data source has odd-numbered years in 1985-2009 and "selected only records representing completed interviews for occupied and vacant units, excluding usual residence elsewhere (URE) and non-interview records." per HADS documentation file. https://www.huduser.gov/portal/datasets/hads/HADS_doc.pdf

Methods

Transformations to achieve normality were conducted on some variables that did not follow a normal distribution. A Log-Log transformation was conducted on OTHERCOST, ZINC2, VALUE, LMED, FMR variables. A Log-Level transformation was conducted on all other variables included in the regression model. Regression model has an adjusted r-square value of 0.52, suggesting the linear regression model explains 52% of the market value variation.

Analysis

Multivariate linear regression analysis was conducted to identify any relationship between variables in the HADS dataset regarding factors that predict or contribute to the current market value of single family units and to model that relationship. Block-wise selection was used to choose predictor variables. The predictor variables were grouped into blocks based on research conducted on contributing factors to resale value of housing unit. Specifically, location, size and layout, condition and age, upgrades, negative events (fires, etc.) show to have an effect on home resale value ("5 Factors That Influence Your Home's Resale Value.", n.d., para. 5-7, 9). Stepwise selection was applied ($\alpha E = 0.15$, $\alpha R = 0.15$) to add variables to the regression model to improve fit.

Variables used for regression model include:

- LN_VALUE_2013: Current market value of unit.
- BEDRMS_2013: number of bedrooms in the housing unit which can reflect size and layout.

- LMED_2013: average median income for the area which can reflect location and income.
- FMR_2013: fair market monthly rate which can reflect cost.
- BUILT_2013: when the housing unit was built which can reflect age and/or condition of housing unit.
- OTHERCOST_2013: other utility costs for the occupant of the housing unit.
- ROOMS_2013: number of rooms in the unit which can reflect size and layout.
- UTILITY _2013: cost of utilities for owning the home.
- REGION_WEST_2013: location in the United States; Western region.
- REGION_NORTHEAST_2013: location in the United States, Northeastern region.
- REGION_MIDWEST_2013: location in the United States, Midwestern region.
- ALL_OTHER_METROS_2013: location to metropolitan city.
- LN_ZINC2_2013: annual household income of the housing unit occupant.
- AGE_2013: Age of head of household.
- PER_2013: Number of persons in household.
- ZADEQ_MODERATELY_INADEQUATE_2013: Adequacy of unit which can reflect any upgrades or negative events that have occurred to the housing unit, moderately inadequate conditions.
- ZADEQ_SEVERELY_INADEQUATE_2013: Adequacy of unit which can reflect any upgrades or negative events that have occurred to the housing unit, severely inadequate conditions.

Regression model listed below:

LN(VALUE) = β0 + β1*BEDRMS + β2*LN(LMED) + β3*LN(FMR) + β4*BUILT + β5*LN(OTHERCOST) + β6*ROOMS + β7*UTILITY + β8*REGION_WEST + β9*REGION_NORTHEAST + β10*REGION_MIDWEST + β11*ALL_OTHER_METROS + β12*LN(ZINC2) + β13*AGE + β14*PER + β15*ZADEQ_MODERATELY_INADEQUATE + β16*ZADEQ_SEVERELY_INADEQUATE

Results

Interpretation of regression model predictor variables and their coefficients in relation to outcome variable (VALUE):

- Intercept: has no practical managerial value.
- β1: one additional bedroom in the housing unit corresponds to a 5.72% decrease in market value of the housing unit all other variables held at their current level.
- β2: a 1% change in area median income is associated with 0.38% increase in market value of the housing unit all other variables held at their current level.
- β3: a 1% change in fair market monthly rent is associated with 0.65% increase in market value of the housing unit all other variables held at their current level.
- β4: every additional year since the housing unit was built is associated with a 0.25% increase in market value of housing unit all other variables held at their current level.

- β5: a 1% increase in the sum of other monthly costs is associated with 0.27% change in market value of the housing unit all other variables held at their current level.
- β6: one additional room in the housing unit corresponds to a 10.40% increase in market value of the housing unit all other variables held at their current level.
- β7: one dollar increase in monthly utility costs corresponds to with 0.04% increase in market value of the housing unit all other variables held at their current level.
- β8: when the housing unit is located in the western region of the US the market value of the housing unit increases by 29.12% as compared to the southern region of the US all other variables held at their current level.
- β9: when the housing unit is located in the northeastern region of the US, the market value of the housing unit increases by 14.58% as compared to the southern region of the US all other variables held at their current level.
- β10: there is no statistical significance for the market value of the housing unit and the housing unit being located in the Midwest.
- β11: when the housing unit is located outside of the "Central City" metropolitan area, the market value of the housing unit increases by 8.38% as compared to the "Central City" metropolitan area all other variables held at their current level.
- β12: a 1% increase in annual household income is associated with a 0.14% increase in market value of the housing unit all other variables held at their current level.
- β13: a year increase in age for the head of household corresponds to 0.22% increase in market value of housing unit all other variables held at their current level.
- β14: with each increase of one person in the household there corresponds to a 2.69% decrease in market value of the housing unit all other variables held at their current level.
- β15: when the housing unit is in moderately inadequate condition there corresponds to a 14.97% decrease in market value of the housing unit as compared to an adequately conditioned housing unit all other variables held at their current level.
- β16: when the housing unit is in severely inadequate condition there corresponds to a 10.23% decrease in market value of the housing unit as compared to an adequately conditioned housing unit.

Regression Statistics								
Multiple R	0.723712232							
R Square	0.523759394							
Adjusted R Square	0.523506881							
Standard Error	0.546714991							
Observations	30193							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	16	9919.483046	619.9676904	2074.183103	0			
Residual	30176	9019.524362	0.298897281					
Total	30192	18939.00741						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-5.730871444	0.381905809	-15.00598135	1.02286E-50	-6.479423099	-4.982319788	-6.479423099	-4.982319788
BEDRMS	-0.057191648	0.006983289	-8.189786711	2.71912E-16	-0.070879193	-0.043504104	-0.070879193	-0.043504104
LN_LMED	0.384045834	0.032556409	11.79632055	4.79568E-32	0.320233885	0.447857782	0.320233885	0.447857782
LN_FMR	0.645287739	0.023606576	27.33508447	1.5571E-162	0.599017845	0.691557633	0.599017845	0.691557633
BUILT	0.002803033	0.000130695	21.44717048	2.7727E-101	0.002546865	0.0030592	0.002546865	0.0030592
LN_OTHERCOST	0.273639233	0.003975814	68.82596966	0	0.265846469	0.281431998	0.265846469	0.281431998
ROOMS	0.104007825	0.00302321	34.40310544	1.8871E-254	0.098082204	0.109933447	0.098082204	0.109933447
UTILITY	0.000407035	2.99942E-05	13.57043677	7.95671E-42	0.000348245	0.000465825	0.000348245	0.000465825
REGION_WEST	0.291236609	0.010839666	26.86767355	3.7179E-157	0.269990401	0.312482816	0.269990401	0.312482816
REGION_NORTHEAST	0.145782817	0.010890463	13.38628244	9.50777E-41	0.124437045	0.167128589	0.124437045	0.167128589
REGION_MIDWEST	-0.016945619	0.009286885	-1.824682698	**0.068058808**	-0.035148309	0.001257072	-0.035148309	0.001257072
ALL_OTHER_METROS	0.083817082	0.007910224	10.59604372	3.45624E-26	0.068312705	0.099321458	0.068312705	0.099321458
LN_ZINC2	0.138542588	0.003512244	39.445602	0	0.131658439	0.145426736	0.131658439	0.145426736
AGE	0.002212675	0.000231092	9.574852946	1.09508E-21	0.001759725	0.002665626	0.001759725	0.002665626
PER	-0.026888764	0.002630883	-10.22043471	1.76081E-24	-0.032045406	-0.021732122	-0.032045406	-0.021732122
ZADEQ_MODERATELY_INADEQUATE	-0.149719223	0.027844908	-5.37689767	7.63425E-08	-0.20429643	-0.095142016	-0.20429643	-0.095142016
ZADEQ_SEVERELY_INADEQUATE	-0.102331419	0.031717495	-3.226339843	0.00125518	-0.164499061	-0.040163777	-0.164499061	-0.040163777

Multiple linear regression table. Red font line shows REGION_MIDWEST dummy variable has p-value greater than alpha 0.05. R Square and Adjusted R Square are highlighted in yellow.

A pairwise correlation test was conducted on all predictor variables in an effort to help detect multicollinearity. No variables are correlated by 80% or more; which can be translated to no variables are highly correlated to each other, with high correlation being a sign of possible multicollinearity.

Conclusion:

There are many factors contribute to the housing or current market value of single family units. Factors of particular interest in contributing to the current market value of single family housing units include the number of rooms in the housing unit, number of bedrooms in the housing unit, location in the western and northeastern region of the US as compared to the southern region, and if the housing unit is in moderately or severely inadequate condition as compared to adequate condition. Future work could include improving the regression model so that it explains more of the current market value variation of the housing unit.

Appendix

How is market value determined in the real estate market?. (n.d.). Retrieved from https://www.investopedia.com/ask/answers/072915/how-market-value-determined-real-estate-market.asp.

5 Factors That Influence Your Home's Resale Value., n.d., Retrieved from https://money.usnews.com/money/personal-finance/articles/2014/07/02/5-factors-that-influence-your-homes-resale-value.