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

**Executive Overview**

The business question analyzed: What predicts 2013 market value of single family housing units? There are many factors that predict the housing or current market value of single family units for 2013. Some of the factors that have statistical significance and seem to contribute are: number of bedrooms in the housing unit, average median income for the area, cost or fair market monthly rental rate of the housing unit, the year in which the housing unit was built, and number of rooms in the housing unit.

**Introduction**

The business question analyzed: What predicts 2013 market value of single family housing units? Linear regression analysis was conducted to identify any relationship between variables within the Housing Affordability Data System dataset regarding factors that predict the current market value of single family housing units for 2013.The conclusion of the analysis was there are many factors that predict the housing or current market value of single family units for 2013. Features that have statistical significance and seem to contribute are:

- BEDRMS 2011: number of bedrooms in the housing unit
- LMED 2011: average median income for the area
- FMR 2011: cost or fair market monthly rental rate of the housing unit
- BUILT 2011: the year in which the housing unit was built
- ROOMS 2011: number of rooms in the housing unit
- REGION 2011: 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
- ZINC2 2011: annual household income of the occupant of the housing unit
- AGE 2011: Age of head of household of the housing unit
- VALUE 2011: Market value of housing unit

**Body**

Question: What predicts 2013 market value of single family housing 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 2011 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 PER_2011, LMED_2011, FMR_2011, BUILT_2011, VALUE_2011, ZINC2_2011, UTILITY_2011, OTHERCOST_2011 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 0.59, suggesting the linear regression model explains 59% 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:

- AGE_2011: Age of head of household.
- BEDRMS 2011: number of bedrooms in the housing unit which can reflect size and layout.
- LN_PER_2011: Number of persons in household.
- REGION_WEST_2011: location in the United States; Western region.
- REGION_MIDWEST: location in the United States, Midwestern region.

- REGION_NORTHEAST: location in the United States, Northeastern region.
- LN_LMED_2011: average median income for the area which can reflect location and income.
- LN_FMR_2011: fair market monthly rate which can reflect cost.
- LN_BUILT_2011: Year the unit was built.
- LN_VALUE_2011: Current market value of unit.
- LN_ZINC2_2011: annual household income of the housing unit occupant.
- ROOMS_2011: number of rooms in the unit which can reflect size and layout.
- ZADEQ_ADEQUACY_2011: Adequacy of unit which can reflect any upgrades or negative events that have occurred to the housing unit, both moderately and severely inadequate conditions.
- ALL_OTHER_METROS_2011: location to metropolitan city.
- LN_UTILITY_2011: cost of utilities for owning the home.
- LN_OTHERCOST_2011: other utility costs for the occupant of the housing unit.

Regression model listed below:

**Ln(VALUE 2013) = β0 + β1AGE + β2BEDRMS + β3Ln(LN_PER_2011) + B4REGION_WEST + B5REGION_MIDWEST + B6REGION_NORTHEAST + β7Ln(LN_LMED_2011) + β8Ln(LN_FMR_2011) + β9Ln(LN_BUILT_2011) + β10Ln(LN_VALUE_2011) + β11Ln(LN_ZINC2_2011) + β12ROOMS_2011 + β13ZADEQ_ADEQUACY + β14ALL_OTHER_METROS + β15Ln(LN_UTILITY_2011) + β16Ln(LN_OTHERCOST_2011)**

*Results*

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

- Intercept: has no practical managerial value
- β1: a year increase in age for the head of household corresponds to 0.086% increase in market value of housing unit, all other variables held at their current level.
- β2: an increase of one bedroom corresponds to 3.588% decrease in market value of the housing unit, all other variables held at their current level.
- β3: there is no statistical significance for the market value of the housing unit and the number of people in the household, all other variables held at their current level.
- β4: when the housing unit is located in the western region of the US the market value of the housing unit increases by 10.026% as compared to the southern region of the US, all other variables held at their current level.
- β5: there is no statistical significance for the market value of the housing unit and the housing unit being located in the Midwest.
- β6: when the housing unit is located in the northeastern region of the US, the market value of the housing unit increases by 3.180% as compared to the southern region of the US all other variables held at their current level.

- β7: each dollar increase in area median income corresponds to a 0.178% increase in market value of the housing unit all other variables held at their current level.
- β8: a 1% change in fair market monthly rent corresponds to a 0.355% increase in market value of the housing unit all other variables held at their current level.
- β9: a 1% change in the year the housing unit was built corresponds to a 3.708% increase in market value of housing unit all other variables held at their current level.
- β10: a 1% change in the current market value of the unit corresponds to a 0.587% change in current market value of the unit, all other variables held at their current level.
- β11: a 1% increase in annual household income is associated with a 0.0468% increase in market value of the housing unit, all other variables held at their current level.
- β12: one additional room in the housing unit corresponds to a 4.969% increase in market value of the housing unit all other variables held at their current level.
- β13: there is no statistical significance for the market value of the housing unit when the housing unit is in moderately inadequate condition or severely inadequate condition compared to adequate condition, all other variables held at their current level.
- β14: when the housing unit is located outside of the "Central City" metropolitan area, the market value of the housing unit increases by 4.256% as compared to the "Central City" metropolitan area, all other variables held at their current level.
- β15: there is no statistical significance for the market value and monthly utility costs for the housing unit, all other variables held at their current level.
- β16: there is no statistical significance for the market value and the sum of other monthly costs for the housing unit, all other variables held at their current level.

| *Regression Statistics* | | |
|---|---|---|
| Multiple R | 0.773180316 | Note: "ALL_OTHER_METROS" is a dummy variable = 1 when METRO3 variable is = '1', otherwise it is =0 |
| R Square | 0.597807802 | Note: "REGION_MIDWEST", "REGION_NORTHEAST" and "REGION_WEST" are dummy variables for the categorical variable REGION |
| Adjusted R Square | 0.597482272 | Note: "ZADEQ_ADEQUACY" is a dummy variable =1 when the categorical variable ZADEQ is ="2" OR "3", otherwise it is =0 |
| Standard Error | 0.489471832 | Note: This regression also takes a logarithmic transformation of many of the 'X' variables |
| Observations | 19785 | Note: This regression also uses the natural logarithm of the 2011 Market Value as one of the 'X' variables |

ANOVA

| | *df* | *SS* | *MS* | *F* | *ignificance F* |
|---|---|---|---|---|---|
| Regression | 16 | 7039.569112 | 439.9730695 | 1836.414387 | 0 |
| Residual | 19768 | 4736.070301 | 0.239582674 | | |
| Total | 19784 | 11775.63941 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | -28.46876578 | 2.208516882 | -12.89044517 | 7.24018E-38 | -32.79764438 | -24.13988718 | -32.79764438 | -24.13988718 |
| AGE 2011 | 0.000868058 | 0.000257736 | 3.368013524 | 0.000758563 | 0.000362874 | 0.001373242 | 0.000362874 | 0.001373242 |
| BEDRMS 2011 | -0.035885372 | 0.007429493 | -4.830123779 | 1.3747E-06 | -0.050447802 | -0.021322941 | -0.050447802 | -0.021322941 |
| LN_PER 2011 | -0.001230281 | 0.007836278 | -0.15699808 | 0.875247965 | -0.016590043 | 0.014129482 | -0.016590043 | 0.014129482 |
| REGION_WEST | 0.100258313 | 0.011159758 | 8.983914661 | 2.84251E-19 | 0.07838425 | 0.122132376 | 0.07838425 | 0.122132376 |
| REGION_MIDWEST | -0.002677998 | 0.010223408 | -0.26194769 | 0.793364501 | -0.022716737 | 0.017360741 | -0.022716737 | 0.017360741 |
| REGION_NORTHEAS | 0.031800174 | 0.011871696 | 2.678654602 | 0.007397974 | 0.008530652 | 0.055069696 | 0.008530652 | 0.055069696 |
| LN_LMED_2011 | 0.178080715 | 0.034186527 | 5.209090524 | 1.91672E-07 | 0.11107225 | 0.24508918 | 0.11107225 | 0.24508918 |
| LN_FMR_2011 | 0.355400737 | 0.023762164 | 14.95658136 | 2.65904E-50 | 0.3088249 | 0.401976575 | 0.3088249 | 0.401976575 |
| LN_BUILT_2011 | 3.708232994 | 0.287226705 | 12.91047428 | 5.59513E-38 | 3.145244525 | 4.271221463 | 3.145244525 | 4.271221463 |
| LN_VALUE_2011 | 0.586730767 | 0.006650353 | 88.22550142 | 0 | 0.573695515 | 0.599766018 | 0.573695515 | 0.599766018 |
| LN_ZINC2_2011 | 0.046791803 | 0.004296218 | 10.89139485 | 1.51704E-27 | 0.038370856 | 0.055212751 | 0.038370856 | 0.055212751 |
| ROOMS 2011 | 0.049687272 | 0.003431601 | 14.47932762 | 2.85776E-47 | 0.042961046 | 0.056413498 | 0.042961046 | 0.056413498 |
| ZADEQ_ADEQUACY | -0.036620853 | 0.023662464 | -1.547634845 | 0.121726245 | -0.08300127 | 0.009759563 | -0.08300127 | 0.009759563 |
| ALL_OTHER_METROS | 0.042565793 | 0.008668048 | 4.910654818 | 9.14971E-07 | 0.02557569 | 0.059555896 | 0.02557569 | 0.059555896 |
| LN_UTILITY_2011 | 3.84903E-05 | 6.6615E-05 | 0.577801714 | 0.563404579 | -9.20808E-05 | 0.000169061 | -9.20808E-05 | 0.000169061 |
| LN_OTHERCOST_201 | 0.010707253 | 0.00561791 | 1.905913896 | 0.056675867 | -0.000304323 | 0.021718829 | -0.000304323 | 0.021718829 |

Multiple linear regression table. Red font line shows PER_2011, REGION_MIDWEST dummy variable, ZADEQ_ADEQUACY, UTILITY 2011, AND OTHERCOST_2011 have p-values 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 that predict the current market value of single family housing units. Factors of particular interest in predicting the current market value of single family housing units include number of bedrooms in the housing unit, average median income for the area, cost or fair market monthly rental rate of the housing unit, the year in which the housing unit was built, and number of rooms in the housing unit. Future work could include improving the regression model so that it predicts with greater accuracy the current market value 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.