Effects of Structural Characteristics on House Prices

Sean Peralta Garcia (23088091)

1 Executive Summary

2 Introduction

Owning a house is one of the biggest investments a person can make. As such, it is of great interest to prospective buyers, sellers and lenders to accurately predict the price of a home. There are a range of models and techniques that attempt to predict the price of a home, from observational appraisals all the way to machine learning models[2]. One method widely studied is that of hedonic pricing. A Hedonic Pricing Model (henceforth, HPM) is a model that attempts to estimate the price of a good by taking its observable characteristics and then weighting them according to their relative impact on the price. Models utilise a range of measures that can be categorised into several groups, namely structural, neighbourhood, and environmental. [3] Hearth and Maier, in a literature review of HPMs for real estate, had identified that the neighbourhood and environmental factors were generally over-researched. While social factors and the "implicit value of structural characteristics" was under-researched.[1]

2.1 Related Work

There appears to be a consensus that creating a HPM, particularly focussing on structural characteristics, has problems with heteroskedasticity. This means that linear models may not be entirely appropriate to estimate the response. This was sought to be relieved by Selim, Limsombunchai and Malpezzi by using a semi-logarithmic form wherein the response variable is transformed by the natural log. [2, 3, 4] Additionally, Malpezzi explains that this effectively allows value added to the house to be proportional to other variables in the model and for an easier interpretation of coefficients such that the coefficient of a measure is the percentage change for 1 unit difference in the measure.[3] In the end Selim concluded that water system, presence of a pool, the type of house, number of rooms, house size, locational characteristics and building type were the most significant variables to affect house prices in the country of Turkey. [4]

2.2 Data Set

The data set contains data on houses in Saratoga County, New York; collected by Candice Corvetti in 2006. It contains 1063 randomly selected observations collecting data on house price, size in square feet, number of bathrooms, number of bedrooms, the presence of a fireplace, land size in acres and the age of the house in years.

2.3 Aim

The aim of this report is to fit a model that will attempt to determine how the price of a house depends its structural characteristics. Significant variables that appear in the Turkish model will be explored in the Saratoga data set to see if they also play a significant role in predicting house price.

3 Methodology

The methodology to be used in this report will be split into two parts. Data exploration and model fitting.

3.1 Data Exploration

Firstly, the scatter plots between independent and dependent variables will be visually analysed to identify relationships and distribution of data. Then the scatter plots between independent variables will be visually analysed to quickly identify multicollinear variables. Additionally, correlation analysis will be used to supplement the visual analysis. This process will aim to identify redundant and non-contributing variables that we can safely assume can be excluded from the final model. Variables that are suspected to have interaction are investigated by running linear regressions with only those interaction terms and analysing the interaction plots. If an interaction effect is apparent then that interaction term will be considered in model fitting.

3.2 Model fitting

Model fitting will be done with backwards elimination, including found interaction variables. Firstly, the collinear variables will be removed from the model and tested using the F-test to check for a better fit. Then normal backwards elimination will be applied.

4 Results

4.1 Data Exploration

- Size - strong, positive linear
- Baths - weak, positive linear
- Bedrooms - weak, maybe non linear (perhaps a parabola)
- Fireplace - positive - houses with a fireplace have a higher range of price and higher mean
- Acres - weak (possibly no relationship), non linear (log?) - most houses under 500 occur between a range of acres - expensive houses are generally on smaller sized plots with 1 outlier (most expensive at 6 acres)
- Age - similar to acres

	Price	Size	Baths	Bedrooms	Fireplace	Acres	Age
Price	1.00	0.77	0.67	0.47	0.41	0.18	-0.26
Size	0.77	1.00	0.74	0.67	0.47	0.22	-0.23
Baths	0.67	0.74	1.00	0.51	0.45	0.13	-0.40
Bedrooms	0.47	0.67	0.51	1.00	0.30	0.15	-0.04
Fireplace	0.41	0.47	0.45	0.30	1.00	0.06	-0.24
Acres	0.18	0.22	0.13	0.15	0.06	1.00	0.01
Age	-0.26	-0.23	-0.40	-0.04	-0.24	0.01	1.00

Now possible interaction terms will be investigated. The following are possible interactions that will be investigated

- Baths:Size - some explanation
- Bedrooms:Size
- Acres:Size
- Fireplace:Age

4.2 Model fitting

First fitted the full model - without interactions - to check residual plot (shows a non constant errors in residual plot) high RSE (52.15) - proof of heteroskedasticity shown in the other studies. and then applied the a natural log function to the reponse RSE as suggested by previous studies a lot - RSE lower (0.279) thus indicating a the response variable follows a log relationship to the independent variables in the model

5 Discussion

References

[1] Herath, S. K. Maier, G. (2010). The hedonic price method in real estate and housing market research. A review of the literature.. Institute for Regional Development and Environment (pp. 1-21). Vienna, Austria: University of Economics and Business.

[2] Limsombunchai, V. (2004). House price prediction: Hedonic price model vs. artificial neural network. New Zealand Agricultural and Resource Economics Society Conference, 25-26 June 2004. Blenheim, New Zealand: New Zealand Agricultural and Resource Economics Society.

[3] Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. Housing economics and public policy, 1, 67-89.

[4] Selim, S. (2008). DETERMINANTS OF HOUSE PRICES IN TURKEY: A HEDONIC REGRESSION MODEL . Dogus Universitesi Dergisi , 9 (1) , 65-76 . Retrieved from