EXAMINING THE IMPACT OF EDUCATION LEVELS ON HOUSE VALUES IN AMERICAN CITIES

Shubha Swamy

ECON 4848: Applied Econometrics

December 13, 2018

Abstract

This paper explores the impact of various education levels on the house values across cities in the United States. To explore this relationship, I use a variety of regression models to analyze the impact of education on house values while also accounting for household income as a control variable. Overall, the results suggest that higher levels of education are related to higher house values. By decomposing the sample into those who possess a college degree and those who don't, the results show that those who have a college degree tend to have higher home values than those who do not.

I. Introduction

What is the effect of education on housing prices in American cities? Various economic factors have the ability to influence housing costs. In this research project, I explore if the difference in education levels affect housing prices in cities across the United States. Housing costs vary significantly across cities in America. Housing prices define its neighborhood characteristics. In this paper, I examine if education influences house values. By analyzing several regression models and statistical tests across variables, I conclude if there is significant evidence to find that education affects housing prices in American cities.

II. The Data

The data for this sample was obtained through IPUMS. The data sample is based on a few primarily variables-- the value of the house, education level attained, total household income, and the city. Since we are interested in measuring the impact of education on housing prices, the dependent variable is the total value of the house. The house value is dependent on the education level variable, which is an independent variable. The sample also includes the total household income variable to account for the control. Another variable which indicates the U.S. city was also included since this research question is limited to cities in the United States.

Before using this data, I had to process the data to get rid of the unusable values and also to generate new variables from existing ones. I removed missing values and produced several new variables for the analysis. I created two new transformed variables which were the log of the total household income and the log of the housing value. I created a new binary variable called *collegegrad* which indicates if the individual has attained a college degree-- this includes an associates degree, a bachelor's degree, or an advanced degree beyond a bachelor's. In addition, I

created several dummy variables which represented mutually exclusive categorical variables for the various levels of education attained.

III. Descriptive Statistics

The primary descriptive statistics are summarized in Table 1 .We conclude that the median price of a home is valued at $315,000 and the average house value is $481,407.70 in cities across the United States. 25% of the houses in American cities are valued below $160,000, and 75% of houses in American cities are valued below $590,000. The data is skewed right resulting in a higher mean than median for home values, caused by the low number of high value homes in our data sample. We also see that the median household income for our data sample is $95,000. 40.84% of individuals in the data set have attained a college degree.

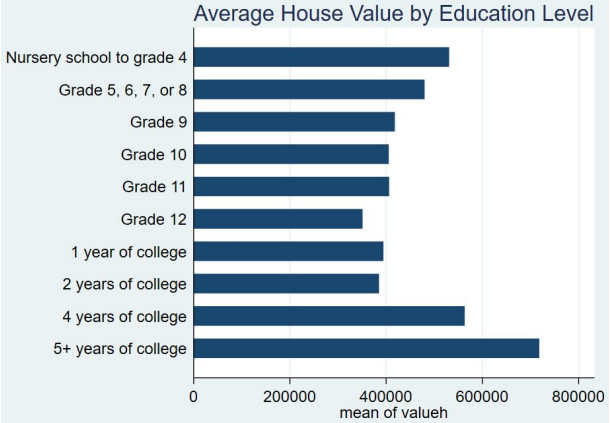
Table 1: Summary Statistics for House Values

Statistics	HH Income (USD)	House Value (USD)	College Degree
N	184472	184472	184472
mean	128946.6	481407.7	0.4084089
25th percentile	54000	160000	0
median	95000	315000	0
75th percentile	155800	590000	1
min	4	1000	0
max	2030000	6308000	1
sd	128533.1	638054.4	0.4915408

The education variable indicates the highest level of education that was completed for an individual. The variable is a categorical variable that stratifies various levels into categories as shown below. Graph 1 displays the average house value for each education level defined in the

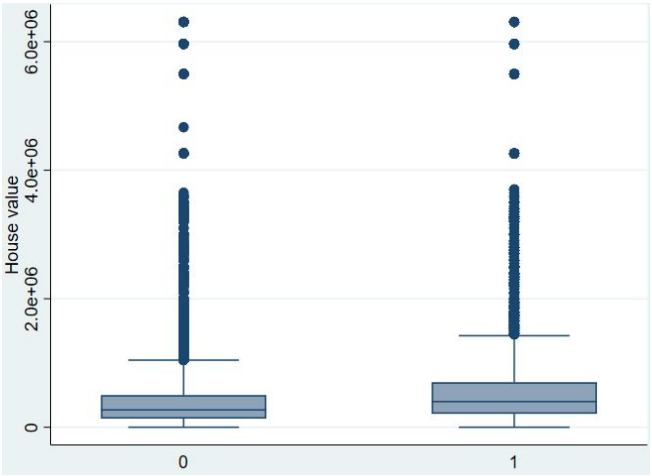
education variable. Those with 5+ years of college have the highest value homes, and those with a high school degree or lower have much lower valued homes on average.

Graph 1: Average Value of House by Education Level



The box plot below portrays the value of the house for individuals for those without a college degree compared to those with a college degree. The value which is classified as '1' represents those with a college degree, anyone with an associate's degree or greater. The value represented as '0' represents those without a college degree, including those who did not finish college with an associates or bachelors degree. As evident in the box plot, we see that the median house value for those with a college degree is higher than the median house value for those who don't possess a college degree.

Graph 3: Box Plot: House Values(USD) for Those With College Degree (1) and Without (0)



IV. Regression Models

In order to conclude the effect of education on house values, I tested several regression models. The first regression examines the direct relationship of the binary variable *collegedeg*, which indicates whether an individual has a college degree or not, with the value of the house. The dependent variable is measured as the logarithm of the house value, and the independent variable indicates 1 for those that hold a college degree and 0 for those who do not. By conducting this regression, we are testing the hypothesis as defined in Equation 2.

$$lvalueh_i = \alpha + \beta * collegedeg_i + \varepsilon_i \quad (1)$$

$$H_o : \beta_{house\ value\ w/\ collegedeg} = \beta_{house\ value\ w/o\ collegedeg} \quad (2)$$

$$H_a : \beta_{house\ value\ w/\ collegedeg} \neq \beta_{house\ value\ w/o\ collegedeg}$$

I included the household income variable as a control variable in the data sample. According to data from the U.S. Bureau of Labor Statistics, income increases with higher education (Vilorio, 2016). Since I examine the relationship between education levels and the value of the housing, it is crucial to control for the income of the household since that could potentially have an impact on the value of the house that can be afforded. Equation 3 examines the relationship between education levels and the value of the house while controlling for the household income. Equation 4 runs a regression test between the household income and housing value equation to confirm if the variable is statistically significant.

$$lvalueh_i = \alpha + \beta_1 * educ_i + \beta_2 * lhhincome_i + \varepsilon_i \quad (3)$$

$$lvalueh_i = \alpha + \beta_1 * lhhincome_i + \varepsilon_i \quad (4)$$

Equation 5 is a multivariate categorical regression equation that examines the impact of various levels of education on house values, controlling for the household income. I generated

several mutually exclusive dummy variables which represent various categories of education. The α in the equation is the base category. In this regression, the α is represented by those who did not attain a high school or GED degree. The regression model defined in Equation 5 predicts the house value for those with the same household income across different levels of education. As shown in Equation 6, we also conduct a joint hypothesis test to check if all levels of education have the same house value, conditional on income.

$$lvalueh_i = \alpha + \beta_1 * hsgrad_i + \beta_2 * collegedrop_i + \beta_3 * assocdeg_i + \beta_4 * bachdeg_i + \beta_5 * advdeg_i + \beta_6 * lhhincome_i + \varepsilon_i \quad (5)$$

$$H_o: \quad (1)\ \beta_{hsgrad} = 0,\ (2)\ \beta_{collegedrop} = 0,\ (3)\ \beta_{assocdeg} = 0,\ (4)\ \beta_{bachdeg} = 0,\ (5)\ \beta_{advdeg} = 0 \quad (6)$$

H_a: Various levels of education do not have the same impact on house values

The final regression (Equation 7) we conduct is an interaction model that tests if those with and without a college degree have homes with the same value from an additional percent increase in household income. We conduct this test by generating an interaction term which is equal to *collegedeg* multiplied by the *lhhincome* variable. The coefficient on this term, β_1, represents the slope to an additional percentage of income for those who hold a college degree and those who do not.

$$lvalueh_i = \alpha_0 + \alpha_1 * collegedeg_i + \beta_0 * lhhincome_i + \beta_1 * collegedeg_i * lhhincome_i + \varepsilon_i \quad (7)$$

V. Results

Table 2: Summary of All Regression Results

Equations:	1	3	4	5	7
	lvalueh	lvalueh	lvalueh	lvalueh	lvalueh
collegedeg	0.444				0.189
	(89.97)**				(2.28)*
educ		0.028			
		(37.43)**			
lhhincome		0.48	0.496	0.449	0.464
		(139.12)**	(144.98)**	(129.27)**	(104.50)**
hsgrad				-0.157	

				(22.44)**	
collegedrop				-0.047	
				(5.70)**	
assocdeg				-0.09	
				(8.59)**	
bachdeg				0.17	
				(24.57)**	
advdeg				0.282	
				(37.29)**	
collegedeg_income					0.005
					-0.65
_cons	12.396	6.919	6.927	7.439	7.191
	(3,713.82)**	(177.52)**	(174.94)**	(185.05)**	(142.19)**
R2	0.04	0.18	0.18	0.2	0.19
N	184,472	184,472	184,472	184,472	184,472
**** p<0.05; ** p<0.01***					

*All regression tests in this project relaxes the assumption of homoscedasticity, and uses the "robust" extension on regressions.

In the first regression equation, we test whether those who hold a college degree and those who don't have the same house value on average. We reject the null, the house value for those who hold a college degree is not the same as for those who do not at the 5% significance level since the p-value on $\beta_{collegedeg}$ is less than 0.05 and the t-statistic value is 89.97, which is high. From this model, we also conclude that having a college degree is associated with 44.4% increase in the house value compared to those who do not have a college degree.

Studies have shown that those in better financial positions can afford higher priced houses (Steegmans, 2017). Higher education levels are also strongly correlated with higher income (Vilorio, 2016). To get an accurate measure of the impact of education levels on housing costs, I included the household income variable as a control variable in subsequent regressions. After running my second regression (Equation 3), the results show that for those with the same

income, an additional level of education is associated with a 2.8% increase in home values. To further understand the impact of income on house values, we examine the direct relationship between income and home values (Equation 4). From the results we see that a 1% increase in income is associated with a 0.496% increase in average house values.

To further understand the impact of each level of education on house prices, I conducted a multivariate categorical variable regression using mutually exclusive dummy variables and controlled for the income. Here we notice an interesting trend. Holding the income for individuals constant-- a high school graduate's house value is associated with a 15.7% less house value compared to a high school dropout's, a college dropout house value is associated with a 4.7% less house value compared to a high school dropout, and an associate degree holder's house value is associated with a 9% less house value compared to a high school dropout. However, we also see that a bachelor degree holder's house value is 17% higher than a high school dropout and an advanced degree holder's house value is associated with a 28.2% higher house value than a high school dropout's house value, holding income constant. To confirm the results, I also ran a joint hypothesis test (Equation 6) which tests the hypothesis that no education level group has different priced house values after controlling for income. I reject the null hypothesis that all education levels have the same house prices, conditional on income at the 5% significance level since the p-value = $0.00 < 0.05$.

The final regression (Equation 7) I performed is an interaction model regression to examine which group (those who hold a college degree and those who don't) have higher value houses, on average, for each additional percentage increase in household income. By running the regression on the interaction term, we fail to reject the hypothesis that college-degree holders and

non-college-degree holders have different returns to a percentage income increase at the 5% significance level. The slope for those with a college degree $(\beta_0 + \beta_1) = 0.469$ is only slightly higher than the slope for those without a college degree $(\beta_0) = 0.464$. At almost all income levels, there is virtually no difference in house values for those with a college degree and those without a college degree.

VI. Biases

Several omitted variables might have potentially impacted my regression models and results in a different manner if they were included. For example, a school district where the house is located in has the ability to influence the regression model in this project. Higher educated individuals might seek to move to better school districts because they might have a higher value for education. The better the school district, the higher the house value for that area. Therefore the omitted variable bias (OVB) on the school district variable is positive. Crime rates can also be correlated to home values and education levels. A higher level of education might mean that education would teach one about the ills of crime, and the likelihood of committing a crime or living in a community with high crime is lower. Similarly, a community with high crime rates can reduce the value of the houses in that neighborhood. Therefore, we can conclude that crime rates have a negative OVB on the regression. Those who have higher levels of education might also seek out to houses that have better utilities and amenities since on average they earn more than those who are less educated. The better the amenities, the higher the house value. Therefore the number of amenities and utilities can have a positive OVB on house values.

VII. Conclusion

Analyzing the impact of education levels on house values through various regression models show that education levels are statistically significant and have a positive impact on house values, especially those who possess college degrees. We see that education levels do have an impact on the house values across cities in the United States, holding household income constant. The general trend in the results showed that the house values were higher for those with a college degree compared to those who did not. However completing only a high school or lower degree negatively affects the value of the house, accounting for income. We can conclude that those who attain a bachelor's, or advanced degree have higher house values. This could be because of the higher level of income they earn. There could also be other variables that we failed to account for as control variables in the regression models that might have an impact on the house values as discussed under the 'Bias' section of the paper. Future extensions of this project can be improved by creating models that account for the variables that were omitted which could potentially have an impact on both the education level and the house value. This study also limited the geographical area to cities in the United States. For future work, one could explore if similar trends hold in rural areas or other regions.

Reference:

Steegmans, Joep. “Financial Position and House Price Determination: An Empirical Study of Income and Wealth Effects.” *NeuroImage*, Academic Press, 6 Mar. 2017, www.sciencedirect.com/science/article/pii/S1051137716301310.

Vilorio, Dennis. “Education Matters : Career Outlook.” *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, Mar. 2016, www.bls.gov/careeroutlook/2016/data-on-display/education-matters.htm.