How are housing prices affected by property tax?

Team 2

12/10/2019

Contents

Introduction

Our team is looking into how housing prices are affected by property tax. We find this question interesting because we want to see if and how property tax is correlated to housing prices. We hypothesize that there is a positive correlation between property tax and housing prices. This would mean that the higher the property tax is the higher the housing costs. We believe this has to do with the location of the house. If a house has a higher property tax does this mean that the house is in a nicer area? Does this mean that the house is in an area with lower crime rates and lower nitric oxide concentrations? Does this mean the the student teacher ratio is lower? These are very interesting questions to be asked as to how property tax correlates with housing prices, in addition to the other factors mentioned.

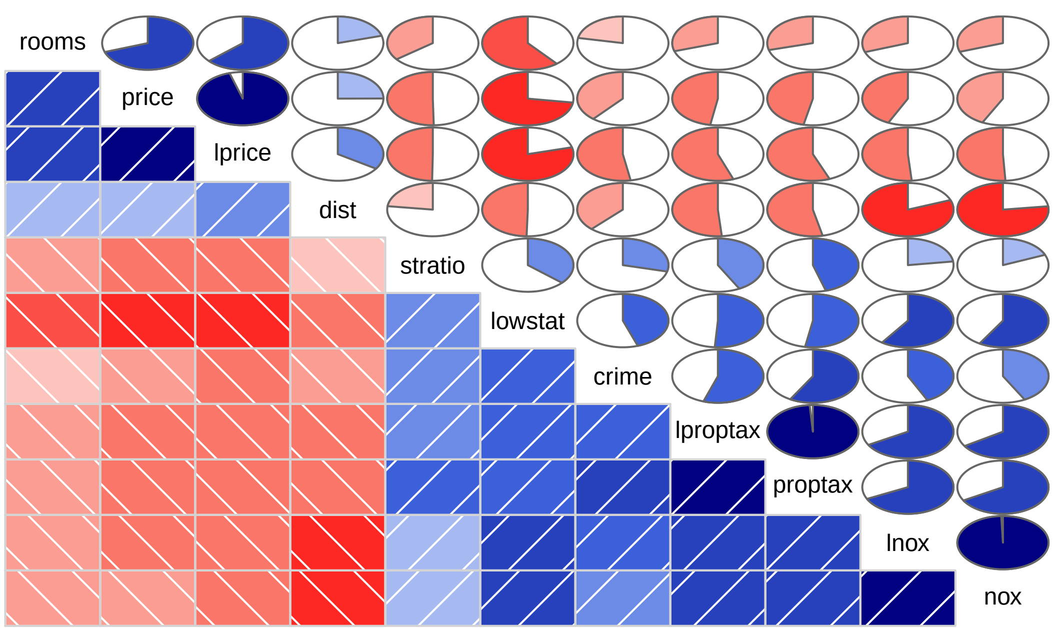
Data

We used the dataset, hprice2, located in the Wooldridge package to answer our above questions. This dataset comes from the paper "Hedonic Housing Prices and the Demand for Clean Air," by Harrison, D. and D.L.Rubinfeld in the Journal of Environmental Economics and Management from 1990. The variables in this dataset are: price (median housing prices), crime (crimes per capita committed), nox (concentration of nitric oxide), rooms (average number of rooms), dist (distance to employee centers), radial (access to highways), proptax (property tax), stratio (student teacher ratio), lowstat (people who are 'lower status'), lprice (log(price)), lnox(log(nox)), and lproptax(log(proptax)). From these variables we chose lprice as our dependent variable and rooms, log(dist), stratio, lowstat, crime, lproptax and lnox as our independent variables. The reason for choosing these variables will be discussed in more detail when we get to the emperical framework section. Below is a summary and correlogram of all of the variables that we have included in our model. In addition to these outputs, we have also included the distribution charts of the variables. These charts help describe the skewedness of our data and can be insightful in determining certain trends such as outliers and range.

Summary of hprice2

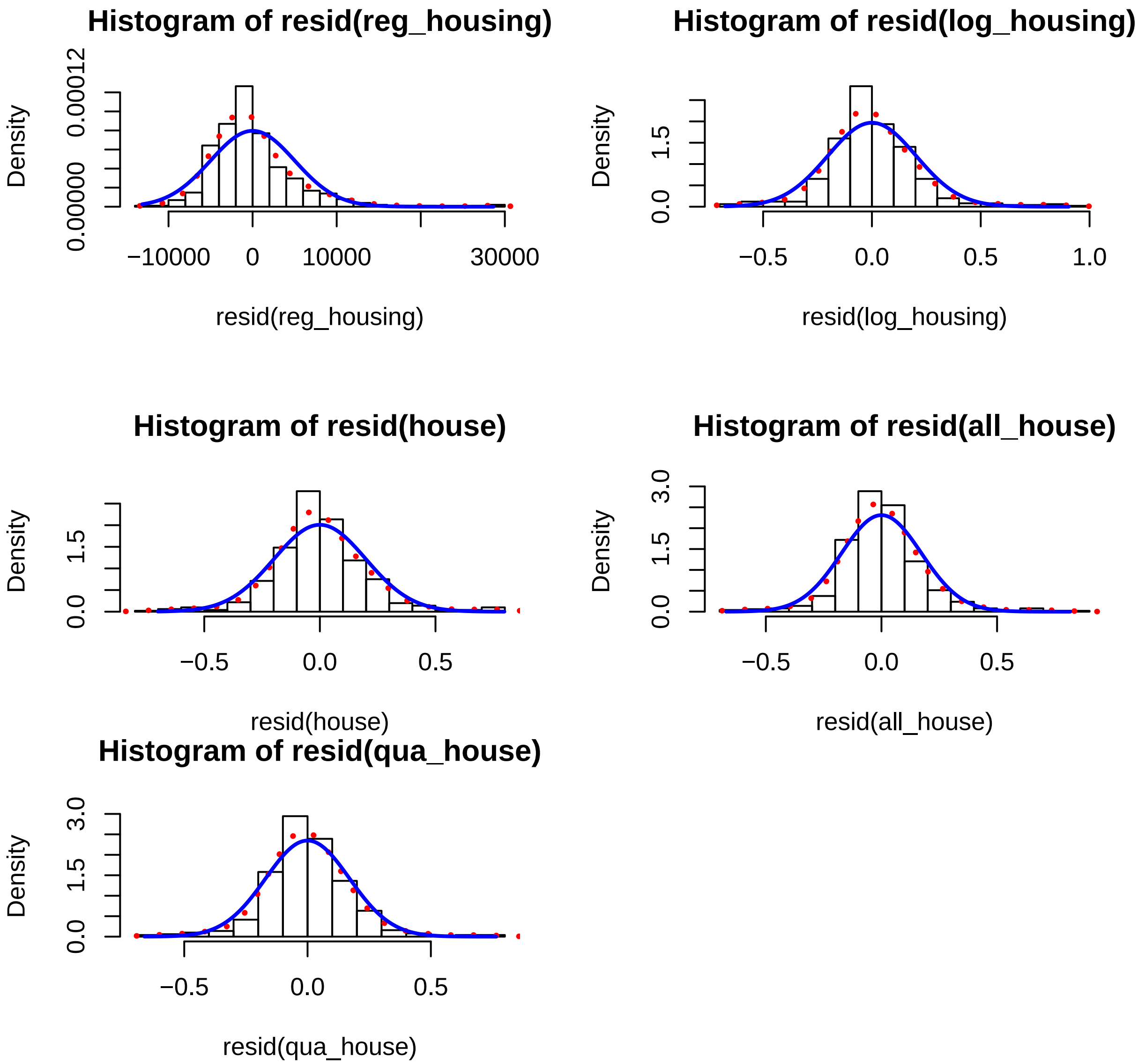
```
     price           crime               nox            rooms      
 Min.   : 5000   Min.   : 0.0060   Min.   :3.85   Min.   :3.560  
 1st Qu.:16850   1st Qu.: 0.0820   1st Qu.:4.49   1st Qu.:5.883  
 Median :21200   Median : 0.2565   Median :5.38   Median :6.210  
 Mean   :22512   Mean   : 3.6115   Mean   :5.55   Mean   :6.284  
 3rd Qu.:24999   3rd Qu.: 3.6770   3rd Qu.:6.24   3rd Qu.:6.620  
 Max.   :50001   Max.   :88.9760   Max.   :8.71   Max.   :8.780  
      dist           proptax         stratio         lowstat      
 Min.   : 1.130   Min.   :18.70   Min.   :12.60   Min.   : 1.730  
 1st Qu.: 2.100   1st Qu.:27.90   1st Qu.:17.40   1st Qu.: 6.923  
 Median : 3.210   Median :33.00   Median :19.10   Median :11.360  
 Mean   : 3.796   Mean   :40.82   Mean   :18.46   Mean   :12.701  
 3rd Qu.: 5.188   3rd Qu.:66.60   3rd Qu.:20.20   3rd Qu.:17.058  
 Max.   :12.130   Max.   :71.10   Max.   :22.00   Max.   :39.070  
     lprice            lnox          lproptax    
 Min.   : 8.517   Min.   :1.348   Min.   :5.231  
 1st Qu.: 9.732   1st Qu.:1.502   1st Qu.:5.631  
 Median : 9.962   Median :1.683   Median :5.799  
 Mean   : 9.941   Mean   :1.693   Mean   :5.931  
 3rd Qu.:10.127   3rd Qu.:1.831   3rd Qu.:6.501  
 Max.   :10.820   Max.   :2.164   Max.   :6.567  
```

Correlations between variables



Histograms of variables

Empirical framework



These are the histograms for each of the regression models. The dotted red line is the density curve and the blue line is the normal curve. This shows that model 4 (all_house) is the closest model to a normal distribution.

```
	studentized Breusch-Pagan test

data:  reg_housing
BP = 43.821, df = 7, p-value = 2.314e-07


	studentized Breusch-Pagan test

data:  log_housing
BP = 59.141, df = 7, p-value = 2.24e-10


	studentized Breusch-Pagan test

data:  house
BP = 80.559, df = 7, p-value = 1.06e-14


	studentized Breusch-Pagan test

data:  all_house
BP = 91.718, df = 10, p-value = 2.443e-15


	studentized Breusch-Pagan test

data:  qua_house
BP = 90.938, df = 11, p-value = 1.092e-14


==========================================================================
                                           price              lprice
                                            (1)                 (2)
--------------------------------------------------------------------------
rooms                           4,416.026*** (416.931)   0.116*** (0.017)
dist                           -1,200.931*** (170.957)  -0.045*** (0.007)
stratio                          -991.786*** (125.692)   -0.036*** (0.005)
lowstat                          -519.460*** (48.510)    -0.028*** (0.002)
crime                             -78.756** (32.986)     -0.009*** (0.001)
proptax                            -3.169 (21.837)        -0.001 (0.001)
nox                            -1,733.523*** (371.126)  -0.070*** (0.015)
Constant                      34,259.480*** (4,883.097) 10.863*** (0.197)
Observations                              506                  506
R2                                       0.704                0.755
Adjusted R2                              0.700                0.751
Residual Std. Error (df = 498)         5,044.938              0.204
F Statistic (df = 7; 498)             169.234***           219.156***
--------------------------------------------------------------------------
Notes:                        ***Significant at the 1 percent level.
                              **Significant at the 5 percent level.
                              *Significant at the 10 percent level.
```

	lprice		
	(1)	(2)	(3)
rooms	0.118*** (0.016)	2.292*** (0.216)	1.168*** (0.352)
log(dist)	-0.260*** (0.033)	-0.188*** (0.030)	-0.177*** (0.030)
I(rooms2)			0.042*** (0.010)
stratio	-0.032*** (0.005)	-0.265*** (0.101)	-0.300*** (0.100)
lowstat	-0.029*** (0.002)	0.201*** (0.021)	0.182*** (0.022)
crime	-0.010*** (0.001)	-0.009*** (0.001)	-0.009*** (0.001)
lproptax	-0.081** (0.035)	1.883*** (0.463)	1.117** (0.495)
lnox	-0.534*** (0.097)	-0.442*** (0.086)	-0.441*** (0.085)
rooms:lproptax		-0.360*** (0.036)	-0.263*** (0.043)
stratio:lproptax		0.042** (0.017)	0.048*** (0.017)
lowstat:lproptax		-0.038*** (0.004)	-0.035*** (0.004)
Constant	11.880*** (0.269)	-0.540 (2.775)	5.739* (3.152)
Observations	506	506	506
R2	0.765	0.822	0.828
Adjusted R2	0.761	0.819	0.824
Residual Std. Error	0.200 (df = 498)	0.174 (df = 495)	0.172 (df = 494)
F Statistic	230.958*** (df = 7; 498)	228.799*** (df = 10; 495)	215.755*** (df = 11; 494)
Notes:	***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.		

Our true version of the equation is:

$$lprice = \beta_0 + \beta_1 * (rooms) + \beta_2 * log(dist) + \beta_3 * (stratio) + \beta_4 * (lowstat) + \beta_5 * (crime) + beta_6 * (lproptax) + \beta_7 * (lnox) +$$

$$\beta_8 * (lproptax : rooms) + \beta_9 * (lproptax : stratio) + \beta_{10} * (lproptax : lowstat) + u$$

Our estimated equation is:

$$\widehat{lprice} = -0.540 + 2.292 * \widehat{rooms} - 0.188 * \widehat{log(dist)} - 0.265 * \widehat{stratio} - 0.201 * \widehat{lowstat} - 0.009 * \widehat{crime} + 1.883 * \widehat{lproptax} - 0.442 * \widehat{lnox} - 0.360 * \widehat{rooms : lproptax} + 0.042 * \widehat{stratio : lproptax} - 0.038 * \widehat{lowstat : lproptax}$$

We decided to use a log model because of the skewedness in our data. We discovered this skewedness from the above distribution histograms. We assumed that a log transformation would help normalize the data and our results showed that we were correct. Using this same method, we decided to use log transformation on some of the explanatory variables which helped with our overall adjusted R^2. Since the above estimated model gave us the highest adjusted R^2 it suggests that this may be the best model for this particular dataset. We tried to incorporate quadratic forms of the variables, but with additional testing these did not help our overall model even though there was statistical significance within model. Even though this seems to be the best model, we still have an issue with heteroskedasticity and thus we have a violation of MLR.5. One way to fix this issue is through transformation, however, with this method we are still in violation of MLR.5. We will use the robust standard error to hopefully help with this issue.

```
t test of coefficients:

                   Estimate Std. Error  t value  Pr(>|t|)    
(Intercept)      -0.5398123  2.7747627  -0.1945   0.84583    
rooms             2.2920934  0.2161074  10.6063 < 2.2e-16 ***
log(dist)        -0.1880624  0.0299578  -6.2776 7.524e-10 ***
stratio          -0.2648742  0.1013421  -2.6137   0.00923 ** 
lowstat           0.2008030  0.0213768   9.3935 < 2.2e-16 ***
crime            -0.0085614  0.0011807  -7.2511 1.603e-12 ***
lproptax          1.8834215  0.4630102   4.0678 5.522e-05 ***
lnox             -0.4422801  0.0861343  -5.1348 4.068e-07 ***
rooms:lproptax   -0.3604126  0.0356131 -10.1202 < 2.2e-16 ***
stratio:lproptax  0.0415093  0.0173187   2.3968   0.01691 *  
lowstat:lproptax -0.0378653  0.0035205 -10.7556 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)    
(Intercept)      -0.5398123  2.5809298 -0.2092  0.834414    
rooms             2.2920934  0.2175222 10.5373 < 2.2e-16 ***
log(dist)        -0.1880624  0.0374701 -5.0190 7.257e-07 ***
stratio          -0.2648742  0.0911166 -2.9070  0.003813 ** 
lowstat           0.2008030  0.0260976  7.6943 7.761e-14 ***
crime            -0.0085614  0.0018037 -4.7465 2.713e-06 ***
lproptax          1.8834215  0.4365764  4.3141 1.936e-05 ***
lnox             -0.4422801  0.1015577 -4.3550 1.618e-05 ***
rooms:lproptax   -0.3604126  0.0370409 -9.7301 < 2.2e-16 ***
stratio:lproptax  0.0415093  0.0157612  2.6336  0.008712 ** 
lowstat:lproptax -0.0378653  0.0045257 -8.3668 6.083e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Linear hypothesis test

Hypothesis:
lproptax = 0

Model 1: restricted model
Model 2: lprice ~ rooms + log(dist) + stratio + lowstat + crime + lproptax + 
    lnox + lproptax:rooms + lproptax:stratio + lproptax:lowstat

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F    Pr(>F)    
1    496                        
2    495  1 18.611 1.936e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on our results from above we can see that using the robust standard error method along with OLS allows us to correct our issue of heteroskedasticity.

```r
vif(all_house)
```

```
            rooms         log(dist)           stratio           lowstat 
       383.064057          4.340379        800.475282        397.788451 
            crime          lproptax              lnox    rooms:lproptax 
         1.709317        559.627252          5.000785        366.788507 
 stratio:lproptax  lowstat:lproptax 
      1499.430891        447.785419 
```

In conclusion, we can say that we have linearity in parameters, homoskedasticity, normality of errors, no perfect collinearity, and we can assume zero conditional mean. In addition, based on our VIF output we can safely assume we do not have issues with multicollinearity. * NOTE: The high VIF for some of the variables is due to the correlation between the variables themselves. We decided to leave these terms in the model because it gave us the highest adjusted R^2 and the variables within the model are siginifcant which allows us to still use this model.

Results

Based on the model above, holding all variables constant except for lproptax, if property tax increases 1 percent from the property tax average then the housing price will decrease by $\approx$ 13.6 percent.This was caluculated with the following equation: $\widehat{lprice} = 1.883 * \widehat{lproptax} - 0.360 * \widehat{rooms : lproptax} + 0.042 * \widehat{stratio : lproptax} - 0.038 * \widehat{lowstat : lproptax}$. We then took the derivative of this equation to obtain:$\widehat{lprice} = 1.883 - 0.360 * \widehat{rooms} + 0.042 * \widehat{stratio} - 0.038 * \widehat{lowstat}$. Plugging in the average of these variables gave us the 13.6 percent. Looking at the above correlogram we can see that lproptax is positively correlated with crime and lowstat, suggesting that if someone is considered lowstat and lives in a high crime area this individiual would have a higher property tax. Given that housing price is inversely correlated to high crime and lowstat (from correlogram) it makes since that higher property tax would be correlated with a lower housing price. Looking at the above output, the R^2 increases as we improve our model because the new model assumes that every independent variable affects the dependent variable, thus this is not a good method to determine the accuracy of our model. If we want to interpret the accuracy of our model we should look at the adjusted R^2. In this case our fourth model seems to be the most accurate. For our t statistics it looks as though each variable is statistically significant in our estimated model. The coefficient of interest, lproptax, is statisitically siginificant at the 1 percent level. (This was calculated by looking at the joint F-statistic of $\beta_6, \beta_8 - \beta_{10}$ which was over 200, suggesting that the variables with lproptax were jointly significant.)

Independent variables partial effect on $\widehat{lprice}$, ceteris paribus:

- if rooms increases 1 unit from the rooms average then lprice increases by 15.7 percent from the average
- if stratio increases 1 unit from stratio average then lprice decreases by 1.6 percent from the average
- if lowstat increases 1 unit from lowstat average then lprice decreases by 2.4 percent from the average

  *NOTE: To calculate these number we use the following method: 1) Take the portion of the equation that holds our variable of interest: $\widehat{lprice} = 2.292 * \widehat{rooms} - 0.360 * \widehat{rooms : lproptax}$ 2) Take the derivative: $\widehat{lprice} = 2.292 - 0.360 * \widehat{lproptax}$ 3) Plug in the average for the variable (i.e. lproptax) 4) Interpret

Independent variables effect on $\widehat{lprice}$, ceteris paribus:

- if log(dist) increases 1 percent lprice decreases 0.18 percent

- if crime increases 1 unit lprice decreases 1.0 percent
- if lnox increases 1 percent lprice decreases 0.442 percent

INTERPRETING INTERACTION TERMS:

Looking at rooms:lproptax we see that a one unit increase in rooms will have a inverse effect on property tax. This makes since as a larger house has a smaller property tax.

Looking at stratio: lproptax we see that a one unit increase in stratio will have a positive effect on property tax. This also makes sense as a larger stratio suggests that a student may go to a larger school. This means that the student may not have as much avilability to the teacher and may not have the resources that a student at a smaller school may have. This suggests that a smaller/cheaper house will have a larger property tax.

Looking at lowstat:lproptax we see that a one unit increase in lowstat will have a inverse effect on property tax. This makes since as a person who is considered low status will most likely have a smaller/cheaper home will have a

SIDE NOTE ON QUADRATIC FORMS: When we tried to incorporate $(log(dist))^2$ we were unable to interpret this variable because when we were looking at the turning point of the variable we did not have β_2 in order to caluculate the turning point value, thus we removed this from the equation even though it was statistically significant.

When looking at the quadratic form of $rooms^2$ we were able to obtain a turning point variable for this function, however, when looking at the effects plot it did not help with our story and was counterintuitive to our conclusion. Since this was the case we removed it from our model.

Our null hypothesis is $H_0 : lproptax = 0$. Based on F statistic results from above we can reject the null hypothesis. This means that our lproptax does not equal zero and is statically significant to our model (p-value = 0).

```
                        2.5 %       97.5 %
(Intercept)      -5.991577319  4.911952658
rooms             1.867492479  2.716694263
log(dist)        -0.246922481 -0.129202267
stratio          -0.463987879 -0.065760511
lowstat           0.158802603  0.242803333
crime            -0.010881224 -0.006241575
lproptax          0.973713771  2.793129209
lnox             -0.611513949 -0.273046241
rooms:lproptax   -0.430384069 -0.290441089
stratio:lproptax  0.007482055  0.075536547
lowstat:lproptax -0.044782371 -0.030948289
```

The above output shows us that our confidence interval does not contain our null hypothesis, meaning we can reject the null hypothesis and conclude that lproptax is statistically significant.

Conclusion

In conclusion, our hypothesis was incorrect. Property tax is inversely correlated to housing price due to contributing factors which are included in the model, such as crime rates, status, and nitric oxide conentration. Higher property tax is positively correlated to low status, nitric oxide concentration, and student teacher ratio. Realistically, this means that those who are disadvantaged by low status, high nitric oxide concentration in their communities, and high student teacher ratios in their schools, also pay higher property tax. In application, this is an example of how the wealth gap between the rich and the poor continues to grow. Those who pay more for their houses, presumably the rich, have a lesser tax burden than those who buy cheaper houses, defined in the model as "low status."