Do Homes Located Near a Metro Station Incur a Price Premium? [1]

Evidence from Washington, DC

Dylan L., Yashwant D., Aru G.

Professor Lo
The George Washington University
dylanucko@gwu.edu
arugupta17@gwmail.gwu.edu
yaswanthsaidevisetti@gmail.com

April 2022

[1] We would like to thank Professor Lo for his help in calculating the haversine distances.

Abstract

This paper investigates

Introduction

Theory

Data

Exploratory Data Analysis

Econometric Specification

This paper utilizes a multiple linear regression approach using ordinary least squares estimators to predict if homes located within a half mile radius of a metro station incur a price premium. It is natural for homes to differ in price for several reasons, including size, year built, geographical location, or number of bedrooms. To determine if the metro station proximity is responsible for the differences in price, this paper controls for many home fixed effects mentioned above and in the data section.

This paper estimates four models, the first three utilize the price of a home as the demendent variable, denoted in dollars. The fourth equation estimates the price premium using the log of the home price to account for skew in the home price data. The general specification of the four models take on the functional form of Equation 1.

$$price = \beta_0 + \delta_1 \; metro.5 + \beta_k \; (house \; fixed \; effects) + u \tag{1}$$

where δ_1 is the paramater of interest and denotes the price differential for a house located within a half mild radius and a house located outside the radius. $metro.5$ is a dummy variable denoting whether a house falls within the radius: $metro.5$=1 if the house lies within the half mile radius and $metro.5$=0 if else.

The term house fixed effects encompasses the home attributes that do not change on average, including the number of bathrooms & half-bathrooms, the lot size, the number rooms, the number of kitchens, and the type of structure.

In Model 4, this paper performs a stepwise selection process to determine which of the nine home fixed effects should be included in the final model. The stepwise selection independently estimates $price$ as a function of the nine home fixed effects and the $metro.5$ variable for a total of ten regressions. The term with the highest R^2 is added to the model, we call this $variable1$. Next, nine regressions are run, each with $variable1$ and the other eight variables; the term with the highest R^2 is added, we call this $variable2$. This process is repeated, adding the variable that results in the highest R^2 and removing variables that are not statistically significant. Model 4, estimates (log)$price$ as a function of the variables chosen by this method.

We expect the sign on $metro.5$ to be positive and moderately large as this paper hypothesizes that homes located within the half mile radius will be more expensive. We also expect the signs on *bedrooms*, *bathrooms*, *size*, *landarea* to all be positive, indicating a higher price given more of these home attributes. We expect some of the structures to have a positive sign and others to have a negative sign, based on the style and quality of the structure. The results of the four regression models can be found in the next section.

To eliminate any multicollinearity issues, this paper examines a correlation matrix to identify independent variables that correlate with one another. Table **??** illustrates the correlation matrix, with high values denoted in red.

Correlation Matrix

	(log)*price*	Price	struc.	metro50	Stories	Land	Bath	Rooms	Half
(log)*price*	1	0.917	-0.074	0.048	0.170	0.158	0.480	0.357	0.321
Price	0.917	1	-0.040	0.031	0.199	0.205	**0.524**	0.389	0.348
sruc.	-0.074	-0.040	1	-0.146	-0.042	**0.505**	0.136	0.095	0.099
metro50	0.048	0.031	-0.146	1	0.014	-0.177	-0.099	-0.170	-0.1050
Stories	0.170	0.199	-0.042	0.014	1	-0.025	0.033	0.030	0.031
Land	0.158	0.205	**0.505**	-0.177	-0.025	1	0.420	**0.505**	0.306
Bath	0.480	**0.524**	0.136	-0.099	0.033	0.420	1	**0.694**	0.290
Rooms	0.357	0.389	0.095	-0.170	0.030	**0.506**	**0.694**	1	0.396
Half	0.321	0.348	0.099	-0.105	0.031	0.306	0.290	0.3963	1

Table 1: Values above 0.50 are highlighted in red. Note that the correlation between log(price) and Price are not considered as they do not appear in the same model.

$$price = \beta_0 + \delta_1\ nearMetro + \beta_1\ medianInc + \beta_2\ crimeRate + \beta_3\ bdrms + \beta_4\ stories + \beta_5\ sqft +$$

$$\beta_6\ fullBath + \beta_7\ halfBath + \beta_8\ grade + \beta_9\ landSize + u$$

where $nearMetro$ is a dummy variable; 1 if within a .5 mile radius from a metro station, $crimeRate$ is percentage of the crime per 100,000, $stories$ is the number of stories a house has, $sqft$ is the square feet of the home, among other home effects variables.

$$price = \beta_0 + \delta_1\ metro.5 + u$$

$metro.5 = 1$ if the house is within a 0.50 mile radius, 0 else

$$\log(price) = \beta_0 + \delta_1 \ metro.5 + \beta_1 \ stories + \beta_2 \ landArea +$$
$$\beta_3 \ cndtn + \beta_4 \ bathroom + \beta_5 \ AC + \beta_6 \ bedroom + \beta_7 \ halfbath + u$$

$metro.5 = 1$ if the house is within a 0.50 mile radius, 0 else

$$price = \beta_0 + \delta_1 \ metro.5 + \beta_k \ (house \ fixed \ effects) + u$$

parameter of interest: δ_1
$metro.5 = 1$ if the house is within a 0.50 mile radius, 0 else

Results

Table 2: Post Treatment Effects on NO_x

	Price		Log(*Price*)	
	(1)	(2)	(3)	(4)
Intercept	677,370.12***	-183455.49	1.2574	1.2903
	(6659.41)	(0.046)	(0.073)	(0.086)
metro50	449324.88***	173696.19	-0.0479	-0.1371
	(57806.31)	(0.064)	(0.078)	(0.100)
bathroom		272713.30	-0.0673	-0.0595
		(0.070)	(0.076)	(0.110)
landarea		47.74	-0.3515**	-0.3334**
		(0.117)	(0.106)	(0.147)
lpop			0.2152***	0.2157***
			(0.044)	(0.044)
rural			-0.1068***	-0.1769
			(0.053)	(0.103)
yr15*rural				-0.016
				(0.154)
ca*rural				0.1801
				(0.122)
yr15*ca*rural				-0.0357
				(0.213)
Fixed Effects	No	No	Yes	Yes
Observations	9778	200	200	200
R^2	0.023	0.119	0.284	0.293
Rdj-R^2	0.022	0.106	0.265	0.264

Notes. Each column reports results from a regression of dummy variables and other indicators for NO_x concentrations. Column (1) uses NO_x concentration in parts per billion. Columns (2) through (4) take the log of NO_x parts per billion concentrations. Fixed effects include variables that do not vary with time such as farm size and farm acerage. The standard errors reported in Table 5 are robust standard errors. HC3 covariance methods are used in the ols model in python ($cov_type =' HC3'$). $***p < .01. ** < .05. *p < .10.$

Analysis

Further Research

Conclusion