The Dual Dynamic: Exploring the Correlation between Housing Prices and Crime Incidence

Dario Tao & Brad Qiang

ECO375: Applied Econometrics

Faculty of Arts and Science, Department of Economics, University of Toronto

Abstract

In a real estate market, homebuyers generally exhibit greater demand for neighborhoods perceived as safer, which should, in theory, drive up housing prices due to increased demand. Conversely, one may hypothesize that affluent neighborhoods could potentially be more susceptible to crimes such as burglary and theft due to the allure of high-value assets.

To clarify the intricate relationship between crime rates and housing prices, we utilize an economic model for our analysis. The findings of our investigation present a nuanced scenario: there exists a moderate correlation indicating that crime rates could have a detrimental effect on housing prices. Nonetheless, the economic significance of this relationship is modest at best. This study invites further exploration into the multifaceted dynamics between safety, crime rates, and their impact on real estate economics.

Context & Data

We obtained the cross-sectional data used in our analysis from D. Harrison and D.L. Rubinfeld's (1978) paper. The bulk of the data was collected by D. Harrison and D.L. Rubinfeld observationally from the 1970 U.S. census, the FBI and the department of education. The data covers 506 census tracts in the Boston Standard Metropolitan Statistical Area in 1970 so the unit of observation is a census tract. The primary variables of interest include natural log of Crimes committed per capita, average student teacher ratio, percentage of people in low status, nitrogen oxide concentration (in parts per 100m), and the average number of rooms in a house. Table 1 provides summary statistics for these variables. Crimes committed per capita ranges from 0.006 to 88.976, suggesting the inclusion of both safe and dangerous census tracts in the dataset; however, the majority of the crimes committed per capita of the census tracts are relatively small with a mean value of 3.61 and standard deviation of 8.59 - this indicates sizable variability in this variable across censuses. Hence, we apply a natural log transformation to this variable to make the distribution of this variable approximately normal. Average student teacher ratio ranges from 12.6 to 22 with a mean value of 18.46 and standard deviation of 2.17 - there is sizable variability in the average student teacher ratios. Percentage of people in low status ranges from 1.73% to 39.07% with a mean value of 12.7% and standard deviation of 7.24% - this is indicative of highly variable levels of this variable across censuses. Nitrogen oxide concentration ranges from 3.85 to 8.71 with a mean value of 5.55 and standard deviation of 1.16 - there is sizable variability in the average student teacher ratios. Average number of rooms in a house ranges from 3.56 to 8.78 with a mean value of 6.28 and standard deviation of 0.70 - there is a moderate variability in the average number of rooms in a house. The main outcome is median housing price in a census tract. The median house price ranges from $5,000 to $50,001 with a mean value of $22511.51 and standard deviation of $9208.86 - the range of median house prices are highly variable in the data.

Regression Analysis

Base Case

To determine the relationship between crimes committed per capita and median house prices, we use a simple linear regression (Table 2, specification (1)) modeled by *median house* $prices_i = \beta_0 + \beta_1 ln(crime)_i + e_i$ for a census tract i in year 1970. Nonlinearities in the data are removed by logging the crimes committed per capita (lncri). Any heteroskedasticity is accounted for by utilizing robust standard errors. Based on our simple linear regression, we find that with a 1% increase in a census tract's crimes committed per capita, there is an associated $19.29 decrease in median house prices, indicating that greater crime rates may be adverse to home prices. The associated negative trend of is visible in Figure 1. The interval between -2,242 and -1,616 is expected to encapsulate 95% of other samples' measure of β_1, which is a $22.42 to $16.16 decrease in median house prices for a 1% increase in a census tracts' crimes committed per capita. Furthermore, the sizable t-value of our test statistic of -12.09 which has a p-value of 0.000. These two tests are indicative that we have adequate evidence to reject the null hypothesis (H_0: $\beta_1 = 0$) that the crime rates does not influence the median house prices; this result is also economically significant given the standard deviation of the *lncri* is 2.16, so a one standard deviation change in this value would be a $4,166 USD change in median house prices, on average. However, this simple linear regression does not carry sufficient evidence to rule a causal effect given the likelihood of lurking variables which would impact the level of crime as well as the median house prices in a census.

To determine if there is a causal effect, we decided to introduce regressors to control for factors which may impact the median house prices. Measurement error? observational error? (from tutorial: measurement error, sample selection/sample bias, omitted variable bias, simultaneous causality, model misspecification).

The two variables we choose to control for in specification (2) is the percentage of people who are considered "lower status," and the average student teacher ratio. The proportion of lower status people is important to control for as it will influence the people's purchasing power which would likely influence the price of homes as well as the affinity to commit crimes. For example, the census tract with a greater percentage of people in a census who are considered "lower status" would have lower income which causes the median house prices to be lower. Based on this finding, it creates a positive bias in our specification (1). Furthermore, there is also positive bias introduced in specification (1) as the census tract with a greater average student teacher ratio would have lower median house prices. In specification (2), the coefficient of determination (R^2) of 0.595 suggests that 59.5% of the variation in median house prices is explained by the three interested variables. To examine the multicollinearity between variables, a correlation matrix is used (Table 3). The correlation coefficient between crime committed per capita and percentage of low status people is 0.61, which is relatively high. However, according to the VIF table, the highest VIF is 1.68, which is relatively low. Collinearity exists when VIF is greater than 5. Therefore, we can conclude that multicollinearity doesn't exist in the model. Mainly, we find that there is an increase in the coefficient of lncri to 260.99 which tells us that lncri has changed from a economically significant negative effect on median house prices in specification (1) to a weak positive effect on median house prices. Now, with a 1% increase in a census tracts' crimes committed per capita becomes associated with $2.60 increase in median house prices when controlling for the percentage of people who are considered "lower status" and the average student teacher ratio. With the t-stat of 1.14 associated with this variable, with an associated probability of 25.6%, which is much greater than the significance level of 5%, we are unable to reject the null hypothesis that there is no effect of crimes committed per capita on median house prices.

We decided to further control for a variety of other variables which are reflective of a census tracts' air pollution levels, and the average home's characteristics. Air pollution levels may be linked to average home prices, as consumers are willing to pay more for clean air, and homebuyers typically pay more for a house with more rooms relative to a house with fewer rooms with other factors held constant. The specific variables controlled for includes nitrogen oxide concentration (in parts per 100 million), and the average number of rooms in a house. These attributes were controlled for in its standalone specification with lncrime in specification (3), as well as together with the percentage of people who are considered "lower status" and the average student teacher ratio in specification (4). In specification (3) we can observe minimal changes relative to specification (2) with respect to lncri in terms of its statistical significance. In specification (3), we find results more similar to the outcomes in specification (1), with the coefficient of lncri possessing a statistically signifiant result with the p-value below 0.05. Therefore, under specification (3), we are able to reject the null hypothesis that the crime rates do not influence the median house prices. However, under specification (4) which controls for all the variables formerly mentioned, we find a statisticially insignificant result for the coefficient of lncri, with a p-value of 0.24. By specification (4), we have insufficient evidence to reject the null that the crime rates do not influence the median house prices.

Extension

Looking beyond the base case, which tested the relationship between median house price and crime committed per capita, we think that crime committed per capita may have a nonlinear relationship with house price. It is because people's willingness to pay for the house may change differently with different safety situations (i.e. different crimes committed per capita**).** For instance, when the crime committed per capita is very high, the census tracts may be considered as very dangerous. For people living in dangerous census tracts, they may feel really insecure and willing to pay much more money to buy a house in safer census tracts. However, when the crime committed per capita is not high, people would not pay much more for houses in a safer census tract. It is because their original census tract is safe enough and there is no willingness to live in a safer census tract. In order to estimate the nonlinear relationships between log crime committed per capita and house prices, we add a squared term into the extension model. We utilize a multiple linear regression modeled by. $\widehat{Price} = \beta_0 + \beta_1 \ln(crime) + \beta_2 \ln{(crime)^2} + \epsilon$. After running the multiple regression model, we find the estimated equation:

$$\widehat{Price} = 21956.82 - 2067\ln(crime) - 200\ln{(crime)^2} + \epsilon$$

In order to test whether the relation is significant, we perform a F test. The F statistic of 97.16 with P value of 0.00 indicates that we can reject the null hypothesis (H_o) that all two variables have no significant effects on house prices. The coefficient of determination (R^2) of 0.2137 suggests 21.37% of the variation in house price is explained by the two variables. We measured the significance of independent variables by using t statistics. Table 3 specification (1) shows that all explanatory variables are statistically significant at 5% level. The coefficient for crime suggests that the average house price decreases by $20.67 for every percent increase in crime committed per capita. The coefficient for crime^2 suggests that the average house price decreases by about $2 for every percent increase in crime committed per capita^2.

In Table 3 specification (2), we are adding one new variable squared lncri to Table2 specification (2). We can see that the omission of lowstat and stratio will result in substantial omitted variable bias. When controlling lowstat and stratio, we can see substantial increases in both lncri and lncri^2 of 2248 and 101 respectively. In specification (2), the coefficient of determination (R^2) of 0.597 suggests that 59.73% of the variation in house price is explained by the four interested variables. The F statistics increase from 97 to 127.07 with a probability of 0.00, which is significant at 0.05 level, indicating we can reject the null hypothesis. However, the effect of lncri and lncri^2 become insignificant.

In Table 3 specification (3), we are adding one new variable squared lncri to Table2 specification (3). We can see that the omission of rooms and nox will result in substantial omitted variable bias. When controlling rooms and nox, we can see substantial increases (1133) in lncri and slight decreases (-36) in lncri^2. In specification (3), the coefficient of determination (R^2) of 0.5617 suggests that 56.17% of the variation in house price is explained by the four interested variables. The F statistics decrease from 97 to 80.09 with a probability of 0.00, which is significant at 0.05 level, indicating we can reject the null hypothesis. According to the t-statistics and p-values, we can see the effects of lncri and lncri^2 on house prices are statistically significant. The effect of log crime rate in the non-linear model is economically significant. The estimated slope coefficient of -934.79 is relatively big. For instance, the standard deviation of lncri in these data is 4902.66, so a one standard deviation change in lncri would be a –45887831 difference in house price. A standard deviation in house price is 9208.83, so a 1 standard deviation difference in lncri is associated with -49949.7% of a standard deviation in house price. In other words, a large difference in median house price is associated with a relatively big difference in lncri.

In specification (4), we are adding one new variable squared lncri to Table2 specification (4). We can see substantial increases (2287) in lncri and slight increases (40) in lncri^2. In specification (4), the coefficient of determination (R^2) of 0.6119 suggests that 61.19% of the variation in house price is explained by the six interested variables. The F statistics decrease from 97 to 90.98 with a probability of 0.00, which is significant at 0.05 level, indicating we can reject the null hypothesis. According to the t-statistics and p-values, we can see the effects of lncri^2 on house prices is statistically significant. The effect of log crime ^2 in the non-linear model is economically significant. The estimated slope coefficient of -266.37 is relatively big. For instance, the standard deviation of lncri in these data is 208.77, so a one standard deviation change in lncri would be a –55610.06 difference in house price. A standard deviation in house price is 9208.83, so a 1 standard deviation difference in lncri is associated with -600% of a standard deviation in house price. In other words, a large difference in median house price is associated with a relatively big difference in lncri^2.

Limitations of Results

One of the limitations of the data is its age. The data was collected from the 1970 U.S. census in Boston, which is quite old. Thus, it may not correctly reflect the true relationship of the house price and crime committed per capita today. House prices across the world and especially America have increased dramatically since 1970 and would not necessarily be reflective of the economic environment in the 21st century. Furthermore, homebuyers' aversion to crimes committed in neighbourhoods may have changed, which in the case of homebuyers becoming more sensitive, would represent a greater delta between censuses with high crime versus those in low crime than what is estimated using the 1970 data.

The next potential limitation is that the model data focuses on the median house price in 506 census tracts in Boston. People in each city are grown up with different cultures, thus their preference of housing would be different. This makes it difficult to apply results to a more general population (i.e. all neighborhoods in the U.S.).

Lastly, the model may not have included all of the omitted variables which were relevant to predicting the median house prices. As the dataset is fully observational data which are simply observed and not obtained experimentally, there is certainly omitted variables not captured. One such potential variable that we can think of, for example, is unrecorded crimes, stemming from the idea that crimes may occur among those with those who are wealthier, however, due to their access to large amounts of money, they can bail themselves out, thereby the crime not being recorded.

Other omitted variables we can think of are the average number of children in households, as more children could mean that the kids are neglected thus leading them down a path of resentment which virtuously leads to behavioural issues and rebellion. Job availability may also be an important attribute, as lower job availability in a census tract may mean leaving many unemployed, thus not being able to pay their bills, leading to crimes such as theft and burglary. Other variables may include religious characteristics, and family conditions with respect to divorce.

Results

This study delves into the link between house prices and a variety of other factors. Our findings reveal a significant, non-linear relationship between the log-transformed per capita crime rates and the median house prices. As crime rates rise within a neighborhood, median house prices are expected to fall.

Our analysis also indicates that several factors — including per capita crime, student-teacher ratio, the percentage of low-income individuals, air quality, and average number of rooms per dwelling — can influence house prices. These factors together explain approximately 67.8% of the variation seen in median house prices.

Thus, it appears that houses in neighborhoods with higher crime rates generally cost less than those in safer areas. While these findings present valuable data for those studying economics, we also acknowledge potential limitations and challenges that may arise.

References

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, *5*(1), 81–102. https://doi.org/10.1016/0095-0696(78)90006-2

Tita, G., Petras, T., & Greenbaum, R. (2006). The Effect of Crime on Property Values: Evidence from a Changing Neighborhood. *Regional Science and Urban Economics*. https://journalistsresource.org/economics/the-impact-of-crime-on-property-values-research-roundup/

Pope, D. G., & Pope, J. C. (2012). The Impact of Crime on Property Values: Evidence from New York City. *The Review of Economics and Statistics*. http://devingpope.com/assets/files/Website_Crime_Property.pdf

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *Journal of Economic Perspectives*. https://pricetheory.uchicago.edu/levitt/Papers/LevittUnderstandingWhyCrime2004.pdf

Gabriel, S. A., & Rosenthal, S. S. (2004). The impact of local amenities on house prices and the quality of life across neighbourhoods. *Real Estate Economics*. https://doi.org/10.1111/1540-6229.12170

Table & Figures

Table 1

Summary Statistics

Variable	Obs	Mean	Std. dev.	Min	Max
price	506	22511.51	9208.856	5000	50001
lncri	506	-.7798935	2.162521	-5.115996	4.488367
rooms	506	6.284051	.7025938	3.56	8.78
stratio	506	18.45929	2.16582	12.6	22
lowstat	506	12.70148	7.238066	1.73	39.07
nox	506	5.549783	1.158395	3.85	8.71
sqrdlncri	506	5.275489	4.494759	3.99e-06	26.17341

Table 2

Base Case

	(1) price	(2) price	(3) price	(4) price
lncri	-1929.0***	261.0	-849.9***	298.2
	(0.000)	(0.256)	(0.000)	(0.238)
lowstat		-838.9***		-530.9***
		(0.000)		(0.000)
stratio		-1216.5***		-1007.4***
		(0.000)		(0.000)
nox			-669.8*	-681.2*
			(0.041)	(0.033)
rooms			7982.0***	4627.9***
			(0.000)	(0.000)
_cons	21007.1***	55825.3***	-24593.0***	22781.6***
	(0.000)	(0.000)	(0.000)	(0.001)
N	506	506	506	506
R-sq	0.205	0.595	0.550	0.673
adj. R-sq	0.204	0.593	0.547	0.670

p-values in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Table 3

Extension

	(1) price	(2) price	(3) price	(4) price
lncri	-2067.0***	181.1	-934.8***	220.4
	(0.000)	(0.424)	(0.000)	(0.386)
sqrdlncri	-200.4*	-98.67	-236.8***	-160.6**
	(0.022)	(0.132)	(0.000)	(0.005)
lowstat		-834.5***		-514.0***
		(0.000)		(0.000)
stratio		-1208.8***		-1003.2***
		(0.000)		(0.000)
nox			-855.0**	-825.0**
			(0.006)	(0.008)
rooms			7982.4***	4710.4***
			(0.000)	(0.000)
_cons	21956.8***	56087.1***	-22384.9***	23555.2***
	(0.000)	(0.000)	(0.000)	(0.001)
N	506	506	506	506
R-sq	0.214	0.597	0.562	0.678
adj. R-sq	0.211	0.594	0.558	0.674

p-values in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Table 4

Pairwise correlation

	lncri	sqrlncri	nox	rooms	stratio	lowstat
lncri	1.0000					
sqrlncri	-0.3313	1.0000				
nox	0.7884	-0.3375	1.0000			
rooms	-0.3070	0.1142	-0.3028	1.0000		
stratio	0.3866	-0.0872	0.1869	-0.3540	1.0000	
lowstat	0.6122	-0.1531	0.5856	-0.6096	0.3654	1.0000

Figure 1

Scatter Plot of Median Housing Price and the Natural Log Transformation of Crimes Committed per Capita