EXPLORING WHY CRIME AND HOUSE PRICES CORRELATE POSITIVELY IN LONDON

Alberto Minghetti[1]
BSc Economics
2nd year
University College London

Explore Econ Undergraduate Research Conference
February 2020

[1] Thanks to Professor Chaudhury, Professor de Paula, Dr Tincani and Lorenzo Incoronato for their advice and guidance.

Abstract

In this paper, I study the relationship between street-level crime rate and property prices in the London area. Initially, OLS estimates suggest a positive correlation between the two, thus contradicting both previous literature on the topic and intuition, which would suggest that crime-ridden and deprived areas are abundant with cheaper housing. In what follows, I suggest that this result is due to issues of endogeneity in the relationship between crime and property prices, and a TSLS estimation approach is thus adopted. After adding appropriate controls to the model, distance from the closest 24-hour police station is used as an instrument that affects house prices only via its effect on crime. This approach suggests that a 1% increase in the local crime rate reduces property prices by roughly 0.7%.

Introduction

The relationship between property prices and crime rates has been widely analysed in the past, and the continuous emergence of new studies on the topic is a testimony of the difficulties that lie behind its estimation using real world data. Determining the sign and the magnitude of this relationship is also relevant for policymakers for at least two reasons. First, it can affect local budgeting decisions regarding crime prevention, as the costs associated with reducing crime in a certain area might be partially offset by increased tax revenues resulting from higher property valuations. Secondly, a policymaker wishing to improve the conditions of a localised group of disadvantaged residents can do so by stepping up its crime reduction efforts in their neighbourhood, which might then lead to asset appreciation and thus increased wellbeing for the group involved.

While there exist microeconomic models based on utility calculations showing that the correlation between the two should be negative, such as (Hellman & Naroff, 1979), policymakers should still look for empirical estimates of the elasticity between crime and property prices to better direct their budgeting decisions. The ability to estimate the sign and magnitude of this relationship using real world data thus becomes relevant and useful. The fact that initial OLS estimates suggest a positive correlation indicates the presence of endogeneity issues, which can be addressed using TSLS estimation.

Empirical Strategy

Widespread crime can reduce property prices through many channels. For instance, (Gibbons, 2004) suggests that a high crime rate might both prevent prospective buyers from considering a certain neighbourhood as well as motivating those who can to move to safer areas, thus reducing local demand for housing and depressing prices. Similarly, a symmetric desire for low crime neighbourhoods also exists, which should thus accentuate the price gap between low and high crime areas.

Furthermore, owning a property in an unsafe neighbourhood might also result in higher insurance and maintenance costs. For instance, (Heeks, et al., 2018) estimate the average costs associated with the occurrence of various types of crime, with an episode of burglary costing almost £6,000. As roughly 20% of the properties in the sample had at least three burglary accidents occurring in their neighbourhood in December, it is easy to see how owning a property in a high-crime neighbourhood can impose significant costs, even in terms of prevention. These risks should in turn be reflected in local property prices. Therefore, a higher occurrence of crimes should correlate with lower house prices.

Nonetheless, issues of reverse causality and general endogeneity associated with the crime rate can complicate the estimation process. For instance, not only do burglaries affect house prices, but also house prices affect the count of burglaries: criminals might wish to maximise their returns by choosing wealthier-looking houses. Similarly, the expected rental returns from letting a property increase with local tourist density due to higher demand for short-term stays, thus affecting prices. However, pick pocketers usually target areas with higher tourist concentration, which in turn increases the crime rate.

In order to avoid these endogeneity problems, I adopt a TSLS estimation approach relying on two assumptions. First, it is assumed that the process of urban regeneration that occurred in the inner London area in recent decades had a significant impact on the crime dynamics of the city[2]. As police stations were mostly built before that period, it is then assumed that their current locations do not necessarily reflect areas of high criminal activity. A testable consequence of this assumption is that crime incidence should increase with the distance from a police station. This is because criminals have now relocated to areas where police have a longer response time.

Once the relationship between crime rates and the distance from the closest police station has been established, a further assumption must be made to use distance as a valid instrument for crime in the property prices equation. In particular, it is assumed that the preference to live close to a police station (if present) can be completely captured by the density of 24-hour police stations within a 1 kilometre radius of the property. Therefore, the distance between a property and the closest 24-hour police station is not correlated to other elements determining house prices and can thus be used effectively as an instrumental variable for crime. In combination with the first assumption, distance from the closest 24-hour police station can now be considered as an exogenous variable influencing both the crime rate and (indirectly) property prices. This approach is an adaptation of what was used in (Gibbons, 2004), where the author applied a similar reasoning to the concentration and distance from pubs.

Given the above assumptions, the first equation of following model can thus be estimated:

$$\ln(price) = \alpha_0 + \alpha_1 \ln(crime) + \alpha_2 nopolstat + \boldsymbol{\alpha^T X} + u_1$$
$$\ln(crime) = \beta_0 + \beta_1 \ln(price) + \beta_2 distpolstat + \beta_3 nopolstat + \boldsymbol{\beta^T X} + u_2$$

[2] The area of King's Cross is a good example. While it used to be synonymous with criminality and deprivation in the 80s and 90s, now it is abundant with offices and expensive housing.

where $\ln(price)$ is the logarithm of property prices, $\ln(crime)$ is the logarithm of nearby crime, $nopolstat$ is the number of 24-hour police stations within 1 km and $distpolstat$ is the distance from the closest 24-hour police station. The controls contained in $\boldsymbol{X}$ will be outlined below. In particular, our assumptions imply that $cov(distpolstat, u_2) = cov(distpolstat, u_1) = 0$ and $\beta_2 > 0$, so $distpolstat$ is both valid and exogenous. Since the dependent variable of the first equation is $\ln(price)$, a hedonic regression will thus be estimated in line with previous literature on the topic[3]. The coefficient in front of each covariate thus indicates how it affects price.

A note of caution should be made. It might be that more crimes are reported in the vicinity of police stations, and thus a reduction in the reported crime rate might occur as we get further away from a station. As no better measures of crime are available, there is little that can be done. However, two facts suggest that underreporting should not be a cause for concern. First, the average distance from a 24-hour police station in London is only around 2km. Secondly, the estimated coefficients in the reduced form equation for crime point to an opposite direction, and the crime rate seems to increase with distance from a station.

DATA

The dataset used for this paper was constructed by "web scraping" pertinent information from every listing referring to the inner London area[4] on property websites using a Python script I wrote specifically for this purpose. For each property the script records the price, the number of bedrooms, the typology of the house, the address, the name of the two closest public transport stops with their type (e.g. underground, DLR, rail) and distance from them. The address is then used to assign coordinates to each observation using GIS software. As less than 1% of the original sample could not be geolocated, this matching mechanism is highly efficient. Note that "price" refers to asking price, and it is assumed that the difference between that and price paid is not relevant.

Data on crime is sourced from the Metropolitan and City of London Police, which make available online anonymised details for each crime recorded, such as typology and coordinates. It is thus possible to use GIS software to assign a measure of local crime incidence to each property. Thanks to the high precision of the data, crime rate is defined as the number of crimes reported in a month[5] within 150 metres of the property. This ensures that neighbourhood-specific crime dynamics are captured.

Data on police stations is again sourced from the Metropolitan and City of London Police website using a purpose-built Python script. The address and opening times for each station

[3] For instance, see (Thaler, 1978).

[4] Here, defined as the area enclosed by the postcodes beginning with N, E, S and W (excluding EN).

[5] The latest available data from December 2019 will be used, since the property prices dataset was built around the same time.

were recorded, and the former was inserted into GIS software to assign coordinates to each observation. It was then possible to count the number of 24-hour police stations within 1 kilometre of the property, and the distance from the closest 24-hour police station for each property in kilometres.

Data on additional controls was sourced from OpenStreetMap. In particular, the number of museums within a 1 kilometre radius of each property is used as a proxy for touristic presence. Similarly, the number of pubs located within 1 square kilometre of each property is an indicator of local amenity density. Amenities play an important role in property valuation, as explained in (Polinsky & Shavell, 1976), and it is thus sensible to include them in the model. In short, the chosen controls are: the number of museums nearby, the number of pubs nearby, dummies for the TfL fare zone, the number of bedrooms, the average distance from the closest public transport stop, whether at least one of the two closest stops belongs to the underground system and the type of property (whether it is a flat, a detached or a terraced house).

Summary statistics for the resulting database are reported in Table 1.

	mean	sd
price	1161464	1805294
number of bedrooms	2.367595	1.241121
flat	.7372573	.4401322
terraced house	.0914104	.2881974
detached house	.0209857	.1433391
pubs nearby	3.355264	3.558538
police stations within 1km	.1638647	.37448
distance from police station	2.284597	1.240698
museums nearby	.7956956	1.559962
ln(price)	13.56103	.7817088
tube	.6794317	.466704
Zone 1	.2356681	.4244238
Zone 2	.3990733	.4897172
Zone 3	.2041895	.4031158
Zone 4	.0696588	.254576
Zone 5	.0045954	.0676348
average distance from stop	.6502947	.3292875
ln(crime_rate)	2.085938	.9831367
Observations	26113	

TABLE 1 - SUMMARY STATISTICS

RESULTS

	(1) ln(price)	(2) ln(price)	(3) ln(price)	(4) ln(crime_rate)
ln(crime_rate)	0.119***	-0.0261***	-0.699***	
	(0.00487)	(0.00405)	(0.173)	
police stations within 1km		-0.0182**	0.0269	0.112***
		(0.00910)	(0.0177)	(0.0168)
distance from police station				0.0277***
				(0.00554)
Controls	No	Yes	Yes	Yes
Observations	26113	26113	26113	26113

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2 - MAIN RESULTS

The main estimation results are presented in Table 2; for the full version, refer to the appendix. As anticipated at the beginning of this discussion, estimating a simple linear model of property prices and crime rate using OLS results in a positive coefficient, which seems to indicate that a 1% increase in the neighbourhood crime rate is associated with a property appreciation of 0.12%, as seen in column 1. A positive result is also found by (Case & Christopher, 1996) and (Lynch & Rasmussen, 2001), with the latter authors explicitly stating its absurdity. In order to get a better estimate, the controls outlined previously are added first, and then distance from a station is used as an exogenous source of variation in crime.

Simply adding controls in column 2 results in OLS estimates already indicating a negative relationship between crime and property prices: a 1% increase in the neighbourhood crime rate is associated with a 0.03% reduction in property prices. While statistically significant at the 99% level, the magnitude of this estimate seems to be too small and not in line with previous literature on the topic, probably due to the issues of simultaneity mentioned above. Therefore, adding a source of exogenous variation in crime should reduce the inconsistency of the estimate.

The source of exogenous variation in the crime rate comes from the distance of a property from the closest 24-hour police station. Column 4 shows the OLS estimates for the reduced form equation for crime, which seem to confirm the urban regeneration assumption: the coefficient on distance is positive, and thus crime increases with distance from a station. However, crime is positively correlated with the density of police stations. This is not necessarily a complete disproval of the urban regeneration assumption: as mentioned previously, it might be that more crimes are reported in the immediate vicinity of a police station.

The TSLS estimates for the first equation of the model are shown in column 3. In line with previous literature on the topic[6], a 1% increase in the neighbourhood crime rate is associated with a 0.7% reduction in property prices. Moreover, the density of 24-hour police station does not have a role in determining prices, as the estimate is positive but not statistically significant.

Conclusion

While these results are interesting, it is sensible to question the solidity of the assumptions involved. In particular, assuming that the presence of police stations only affects house prices via their local density might seem unrealistic. However, it should be noted that the coefficient on station density in column 3 is not statistically significant despite the large sample size, thus suggesting that the presence of police stations does not affect prices at all.

Further research should consider the adoption of a crime index rather than raw data. Not all crimes impose the same costs on houseowners, and it might thus be sensible to weight each crime occurrence by its average cost, as outlined in (Heeks, et al., 2018). Moreover, there is scope for improving the 1 kilometre and 150 metres cut-off rules, as they were chosen solely on the basis of simplicity and intuition. Finally, it might be interesting to analyse the effect of police stations' opening times, thus including in our sample those police stations that are not open 24 hours.

[6] (Hellman & Naroff, 1979) find an elasticity of -0.63. However, their crime rate is defined in terms of population, rather than in terms of area.

APPENDIX

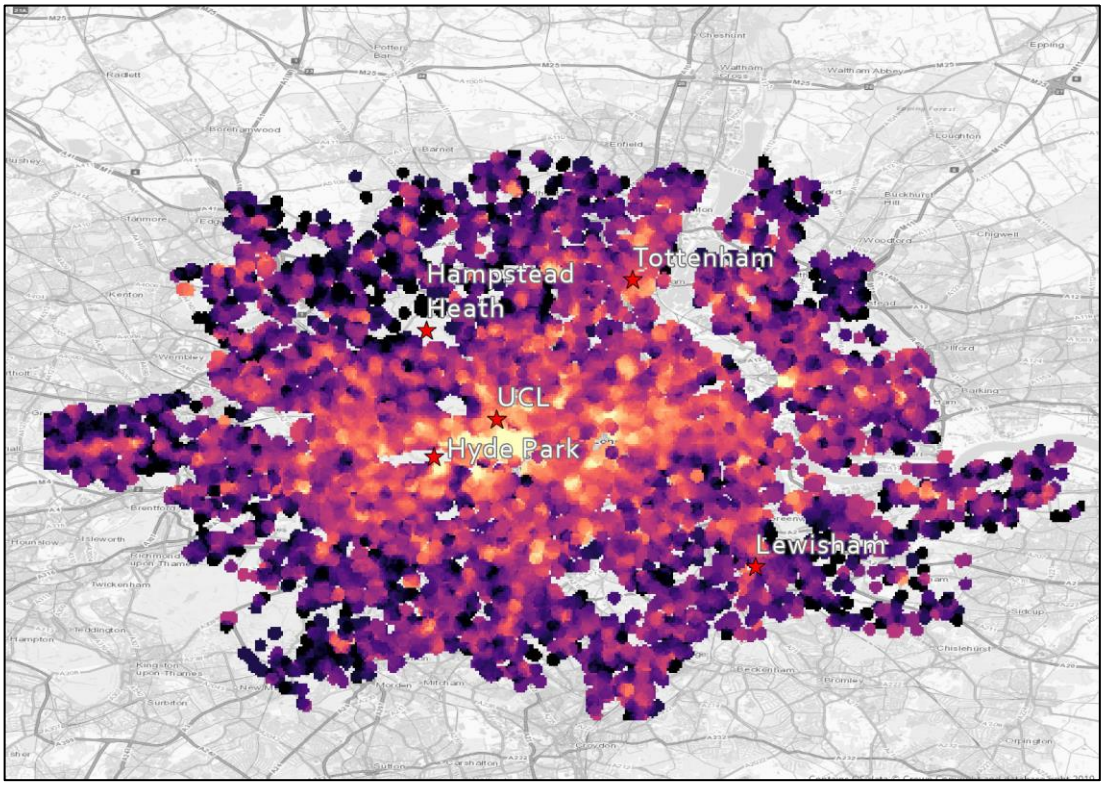

FIGURE 1 - CRIME DENSITY; LIGHTER MEANS MORE CRIME

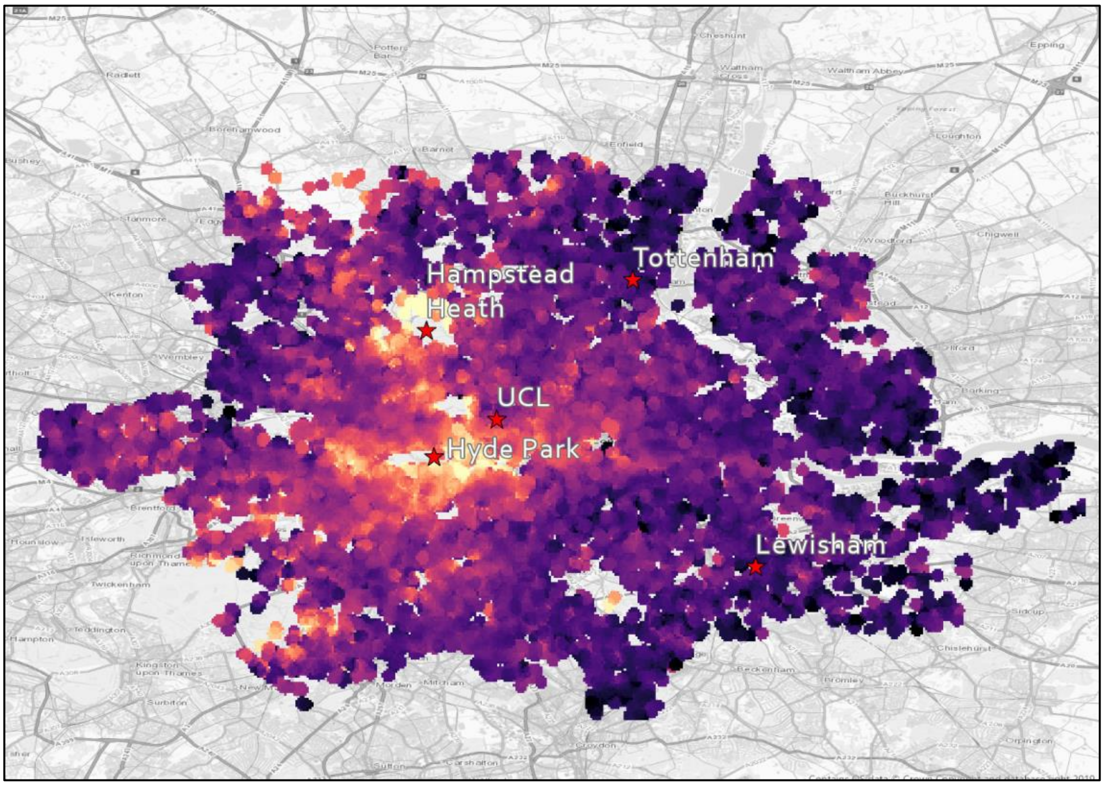

FIGURE 2 - HOUSE PRICES; LIGHTER MEANS MORE EXPENSIVE

	(1)	(2)	(3)	(4)
	ln(price)	ln(price)	ln(price)	ln(crime_rate)
ln(crime_rate)	0.119***	-0.0261***	-0.699***	
	(0.00487)	(0.00405)	(0.173)	
number of bedrooms		0.361***	0.337***	-0.0363***
		(0.00517)	(0.00839)	(0.00539)
average distance from stop		-0.224***	-0.596***	-0.547***
		(0.0111)	(0.0970)	(0.0182)
detached house		0.378***	0.222***	-0.234***
		(0.0281)	(0.0556)	(0.0387)
terraced house		-0.00194	0.0634**	0.100***
		(0.0128)	(0.0252)	(0.0208)
flat		0.0526***	0.147***	0.139***
		(0.0121)	(0.0301)	(0.0166)
Zone 1		0.665***	1.093***	0.659***
		(0.0176)	(0.112)	(0.0262)
Zone 2		0.198***	0.401***	0.321***
		(0.0129)	(0.0548)	(0.0220)
Zone 3		-0.197***	-0.117***	0.125***
		(0.0125)	(0.0279)	(0.0228)
Zone 4		-0.293***	-0.270***	0.0282
		(0.0131)	(0.0230)	(0.0270)
Zone 5		-0.240***	-0.345***	-0.146**
		(0.0324)	(0.0616)	(0.0672)
tube		0.257***	0.233***	-0.0377***
		(0.00723)	(0.0124)	(0.0124)
museums nearby		0.0426***	0.0685***	0.0386***
		(0.00348)	(0.00831)	(0.00441)
pubs nearby		0.000310	0.0395***	0.0603***
		(0.00127)	(0.0102)	(0.00190)
police stations within 1km		-0.0182**	0.0269	0.112***
		(0.00910)	(0.0177)	(0.0168)
distance from police station				0.0277***
				(0.00554)
Constant	13.31***	12.48***	13.77***	1.821***
	(0.0112)	(0.0241)	(0.332)	(0.0368)
Observations	26113	26113	26113	26113

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 1 - FULL REGRESSION RESULTS

References

Case, K. E. & Christopher, M. J., 1996. Housing Price Dynamics Within a Metropolitan Area. *Regional Science & Urban Economics,* 26(3-4), pp. 387-408.

Gibbons, S., 2004. The Costs of Urban Property Crime. *The Economic Journal,* Volume 114, pp. F441-F463.

Heeks, M., Reed, S., Tafsiri, M. & Prince, S., 2018. *Home Office Research Report 99: The economic and social costs of crime, 2nd edition,* London: Home Office Research Development and Statistics Directorate.

Hellman, D. A. & Naroff, J. L., 1979. The Impact of Crime on Urban Residential Property Values. *Urban Studies,* Volume 16, pp. 105-112.

Lynch, A. K. & Rasmussen, D. W., 2001. Measuring the impact of crime on house prices. *Applied Economics,* 33(15), pp. 1981-1989.

Polinsky, A. M. & Shavell, S., 1976. Amenities and property values in a model of an urban area. *Journal of Public Economics,* 5(1-2), pp. 119-129.

Thaler, R., 1978. A Note on the Value of Crime Control: Evidence from the Property Market. *Journal of Urban Economics,* Volume 5, pp. 137-145.