Online appendix for: Expected Returns to Crime and Crime Location

Nils Braakmann
Newcastle University

Arnaud Chevalier
Royal Holloway and IZA

Tanya Wilson
University of Glasgow and IZA

December 4, 2023

Nils Braakmann: Newcastle University Business School, 5 Barrack Road, Newcastle upon Tyne, NE1 4SE.
Email: nils.braakmann@ncl.ac.uk
Arnaud Chevalier: Department of Economics, Royal Holloway University of London, Egham, TW20 0EX.
Email: arnaud.chevalier@rhul.ac.uk
Tanya Wilson: Adam Smith Business School, University of Glasgow, Glasgow, G12 8QQ.
Email: tanya.wilson@glasgow.ac.uk

1 Data

1.1 Street-level crime data

The street-level crime data is available from https://data.police.uk/ for England and Wales (and Northern Ireland). Since December 2010, the available data contains details of each reported crime by month and police force area. Importantly, the data includes a measure of the location of each crime. To minimise privacy risks, the publicly available data adds noise to the location; i.e. the exact location of crime is approximated to the nearest "map point".[1] In the vast majority of cases in urban areas, this largely moves the crime from the exact address where it was committed to the centre of the respective street or from the exact location within a park to the centre of the park. If the nearest map point is more than 20km away, which can occur in rural areas, the co-ordinates are zeroed out. Importantly, the anonymisation process preserves information about the Lower Layer Super Output Area (LSOA) where the crime was committed, which we use as the location of crime.[2]

The data distinguishes between the following crimes: "anti-social behaviour", "burglary", "criminal damage and arson", "drugs", "other theft", "other crime", "public disorder and weapons", "robbery", "shoplifting", "vehicle crime" and "violent crime". Not all of the categories are measured consistently over the observation period. "Criminal damage and arson", "drugs", "other theft" and "shoplifting" have been separate categories only since mid-2011 and were part of "other crime" before. "Public disorder and weapons" also underwent several changes - until mid-2011 it was part of "other crime" and since mid 2013 it has been split up into two separate categories "public disorder" and "possession of weapons". Theft also undergoes several changes with "bicycle theft" and "theft from the person" being split from "other theft" towards the end of 2013. For burglary, robbery, vehicle crime and violent crime, we aggregate at the LSOA level the number of reported crimes per month.

Separate files also contain information on stop and search activities, on a monthly basis and for each police force. This data is available from January 2015, but not all police forces reported stop and search activities by then. When available the data contains exact location of the stop and search, characteristics of the "searched person", ethnicity of the searching officer, legislation under which the search was conducted, and outcome of the search. Altogether, we have information for 1,771,935 searches, but after excluding searches conducted by the British Transport Police (essentially stops and seaches conducted on trains and train stations) and those for which no location information is available, we are left with 1,265,032, or 71% of the total. The location is the exact longitude and latitude, which is then mapped to postcodes based on the nearest distance. To achieve this, we use data from National Statistics Postcode Lookup (https://geoportal.statistics.gov.uk/datasets/1951e70c3cc3483c9e643902d858355b) and the command *geonear* in Stata (Picard, 2010). Postcode information is then used to identify the LSOAs in which the search took place. 99% of locations are matched to a postcode centroid located less

[1] Map points are located over the centre point of a street, or "features" such as commercial premises. Map points must contains at least eight postal addresses or no postal addresses at all.

[2] Three LSOAs do not have any crime reported during the sample period, in each case the closest street and map point is outside the LSOA boundary. The areas are Islington 006F containing the stadium and grounds of Arsenal Football club, Lambeth 024E encompassing a school and park, and Tower Hamlets 031F which contains mainly commercial property adjacent to the River Thames.

than 300 metres away. This process identifies 31,738 LSOAs with at least one stop and search activity recorded, or 91% of the LSOAs used in the main analysis. Over the period, almost 1.8 search and stop activities per LSOA and month were recorded, but there is a large amount of variations over time and between police forces. The median LSOA has 0 stops and searches in a given month but some have more than hundred.

1.2 Lower Layer Super Output Area

LSOAs are an aggregation of adjacent Census Output Areas (OA) with similar social characteristics that align with local authority district boundaries. These OAs were built following the 2001 Census outputs from clusters of adjacent postcode units, and designed to be socially homogeneous (in terms of dwelling types and housing tenure) and of similar population sizes. The OAs tend to follow natural boundaries, such as roads. The OAs target size is 125 households, and cannot be lower than 40, with an average population of 297. The total numbers of OAs in England and Wales in 2011 were 171,372 and 10,036 respectively.

Following the 2001 census, LSOA were created by aggregating four to six OAs so that they have a population between 1,000 and 3,000, and are as homogenous as possible. In 2011, after some minor changes, there were 32,844 LSOAs in England and 1,909 in Wales.

1.3 2011 Census

The 2011 census was conducted in April. Aggregated data at the level of various geographies is available from NOMIS (https://www.nomisweb.co.uk/census/2011). We extract information at the LSOA level on population size, and population by ethnicity, so that we compute the share of the population for each ethnic group. We define South Asian as individuals reporting their ethnicity as Indian, Pakistani or Bangladeshi. Similarly the Black population aggregates the share of individuals identifying as Caribbean black, African black or other black. The social status classification is derived from census data by the Office for National Statistics and splits the UK population into 6 groups based on the employment of the household head : A "Higher managerial, administrative or professional" (often labelled "upper middle class"), B "Intermediate managerial, administrative or professional" ("middle class"), C1 "Supervisory or clerical and junior managerial, administrative or professional" ("lower middle class") and C2 "Skilled manual workers" ("skilled working class"), D "Semi-skilled and unskilled manual workers", also often labelled "working class", and E "Casual or lowest grade workers, pensioners, and others who depend on the welfare state for their income". The model uses occupation, employment status, qualification, tenure, full-time or part-time status to define social grade.

1.4 Other data

The quarterly unemployment rate at the local authority level is based on a model developed by the Office of National Statistics and is available from NOMIS (https://www.nomisweb.co.uk/datasets/umb). For details about the methodology see ONS (2006). The Economic Policy Uncertainty Index is available for the UK from https://www.policyuncertainty.com/. The index was developed by Baker, Bloom and Davis (2016). The index is based on the count of newspaper articles including

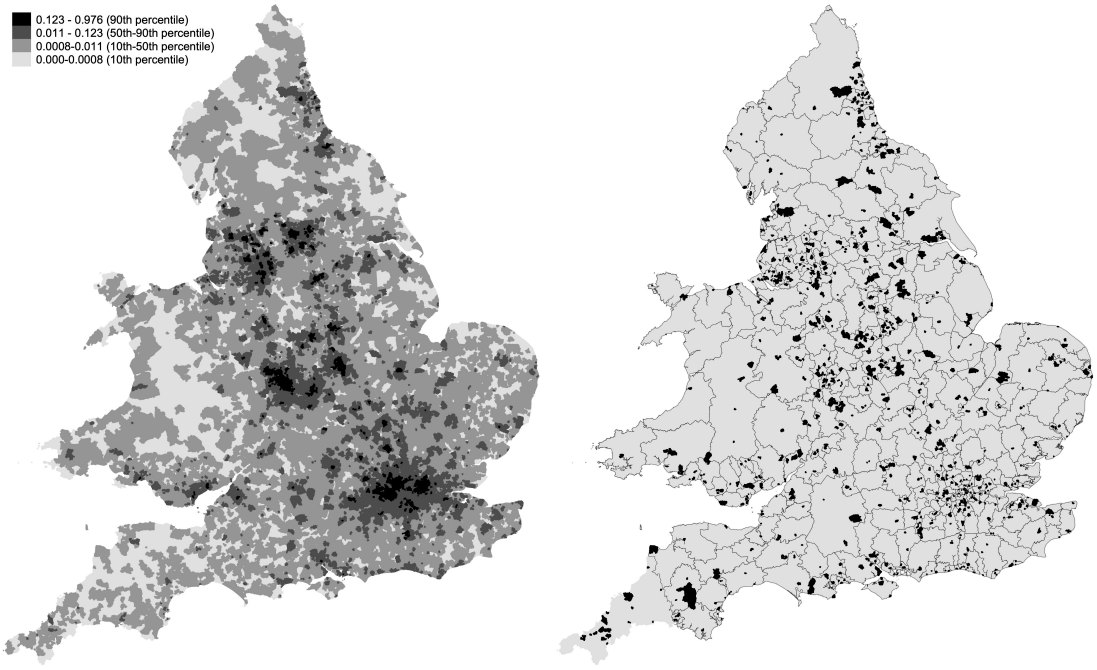
the terms 'policy', 'tax', 'spending', 'regulation', 'Bank of England', 'budget', and 'deficit'. The index is based on 11 British newspapers and updated monthly.

2 Definition of Treatment Areas

We use the 2011 census to determine the share of South Asian households per neighbourhood. Most neighbourhoods have a low share of South Asians, the median is 1.1% but there is substantial heterogeneity with London, the Midlands and North West having neighbourhoods with very high concentrations - see Figure 1, Panel (a). It is clear that defining treatment areas according to a simple national threshold, e.g., a neighbourhood with a South Asian share of household above the 90th percentile for England and Wales would mainly just pick up large urban conurbation effects, as these neighbourhoods are mainly concentrated in the large cities such as London, Birmingham and Manchester. Instead we define treatment areas as neighbourhoods (LSOA) with a high proportion of South Asian households relative to the share of South Asians in their local authority. The advantage of this approach is that it ensures a geographic spread of treatment areas across England and Wales and is consistent with Vandeviver and Bernasco (2019) and Kirchmaier, Langella, and Manning (2021) who find that criminals typically choose targets in the close vicinity of their residence.

Our prime definition of a treatment area is a neighbourhood with a share that is an outlier for the Local Authority, i.e., a South Asian share of households in excess of the 75th percentile plus 1.5 times the interquartile range of all neighbourhoods within a local authority. This defines 2,247 treatment areas which by definition are spread throughout the country (see Panel b). These neighbourhoods have between 0.5% and 97.6% share of South Asian households, with a mean of 16.2% and comprise 6.5% of neighbourhoods in our sample.

Figure 1: Area definitions across England and Wales



(a) Share of South Asian Households (b) Preferred treatment definition

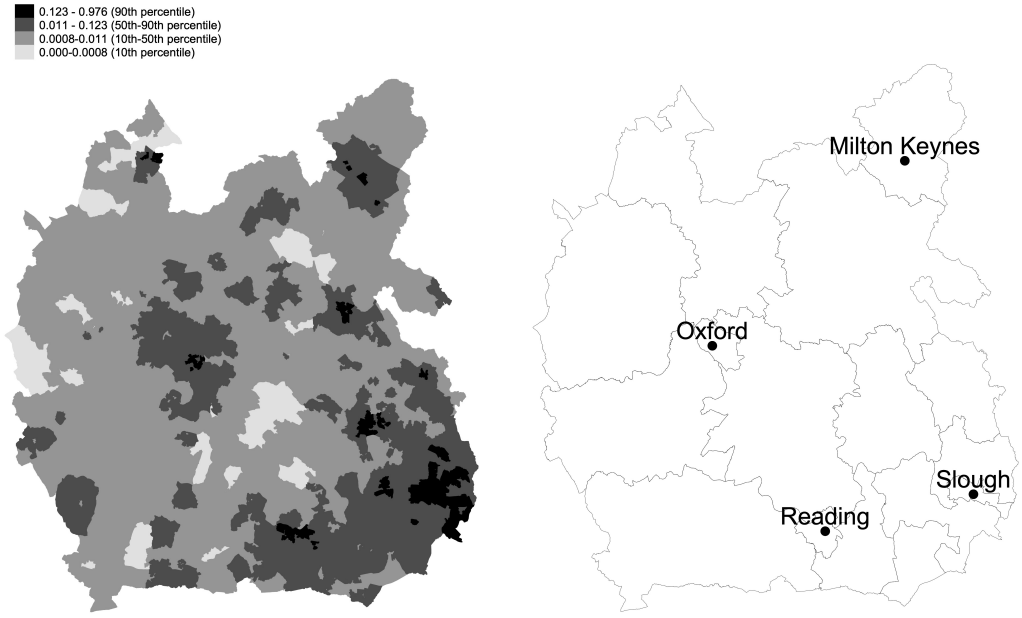
Notes: The maps display the share of South Asian households in each LSOA in England and Wales as enumerated in the 2011 census, and the treatment areas using our preferred treatment definition. Maps are drawn using shp2dta (Crow, 2006) and spmap (Pisati, 2018).

In Table 2 of the paper we also consider alternative treatment definitions. First (Table 2, column 6) we define outlier neighbourhoods as above, but impose a lower bound, such that treated neighbourhoods contain a South Asian share of at least 5%. This more conservative definition accounts for 1,311 (3.76%) neighbourhoods where the share of South Asians ranges between 5.02% and 97.6% and the mean is 26.06%. Our second alternative definition (Table 2, column 7) considers treated neighbourhoods as having a South Asian share of population in excess of the 90th percentile of neighbourhoods within their local authority. This leads to around 10% (3,627) of neighbourhoods being defined as treated. These have a share of South Asian households between 0.2% and 97.6% and mean of 15.9%. Our final alternative definition (Table 2, column 8) instead defines treated neighbourhoods as those in the upper decile of the national distribution. This definition includes 10% of neighbourhoods (3,475), with a South Asian share of households ranging between 12.4% and 97.6% and a mean of 31.4%.

The rationale behind our approach to defining treatment areas becomes even clearer when we consider individual police force areas, and the distribution of South Asian households within each police force area as compared with the national distribution. To illustrate we consider the Thames Valley Police Force Area illustrated in Figure 2, which is the largest non-metropolitan police force area in England and Wales covering an area of over 2,000 square miles and serving a population of more than 2 million people. Figure 2 (a) shows the distribution of South Asian households across the area. The variation in the South Asian population share is lower than nationwide, and ranges between zero and 63.7%, but the median share is higher, at 2.91%. As

can be seen the largest concentrations of South Asian households are close to the four major urban conurbations (population size of over 100,000) in Thames Valley - Oxford in the center, Milton Keynes in the north east, Slough in the south east, and Reading in the south of the police force area, see figure 2 (b) .

Figure 2: South Asian distribution across Thames Valley Police Force Area

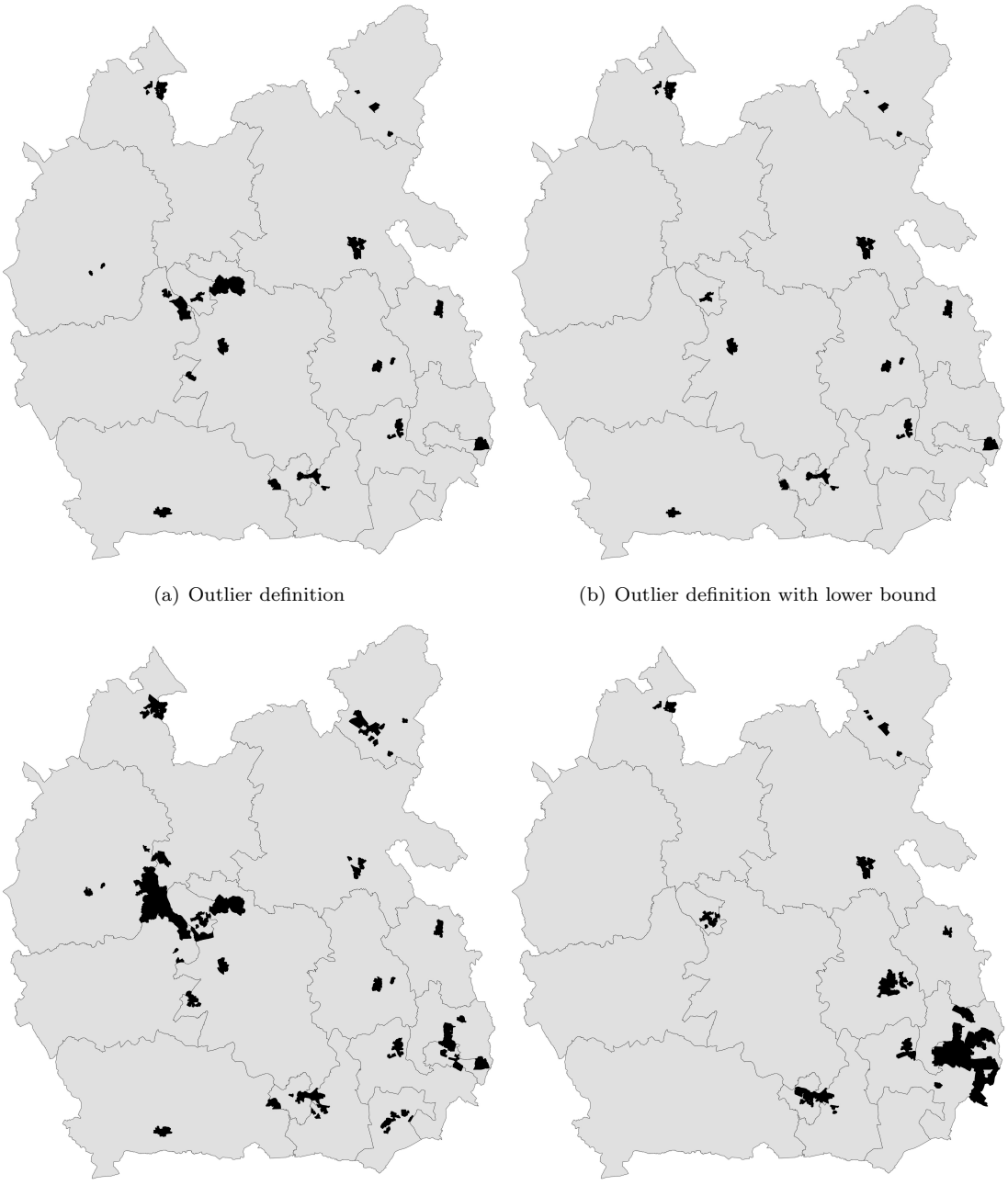


(a) Share of South Asian Households (b) Major Urban Centers

Notes: The maps display (Panel (a) the share of South Asian households in each LSOA within the Thames Valley Force Area as enumerated in the 2011 census; panel (b) shows the four large urban areas and 14 local authorities within the police force area.

Figure 3 compares the four definitions of treated area in the Thames Valley Police Force Area. Our preferred specification where treated areas are defined as an outlier relative to other neighbourhoods in the same local authority is presented in panel (a), which shows all but one local authority has a treated area according to this definition. Panel (b) illustrates the more conservative outlier definition, which reduces the number of neighbourhoods defined as treated particularly in the vicinity of Oxford, and results in a further two local authorities without any treated neighbourhoods. In panel (c) we show treated areas defined as being in the top decile of neighbourhoods in the local authority by South Asian share. By definition this measure assigns at least one treated area in each local authority, and we can see specifically a substantial increase in the number of treated neighbourhoods around Oxford. Panel (d) highlights when treated areas are defined relative to the nationwide distribution this results in a pull towards treated areas being mainly defined in the major urban conurbations (in particular Slough in the south east).

Figure 3: Area definitions across Thames Valley Police Force Area



(a) Outlier definition

(b) Outlier definition with lower bound

(c) Top decile relative to Local Authority

(d) Top decile relative to England and Wales

Notes: The maps display the share of South Asian households in each LSOA within the Thames Valley Force Area as enumerated in the 2011 census, and the treatment areas used in the analysis.

3 Additional robustness checks

In this section we present additional robustness checks omitted from the main paper.

3.1 Statistical inference

We explore the robustness of our results to different clustering levels in Table 3.1. As our panel is formed from observations at the neighbourhood (LSOA) level repeatedly over time, in our main analysis we report estimates clustered at the LSOA level (indicated in bold). This involves clustering using the 34,753 LSOAs.

Table 1: Robustness to different levels of clustering

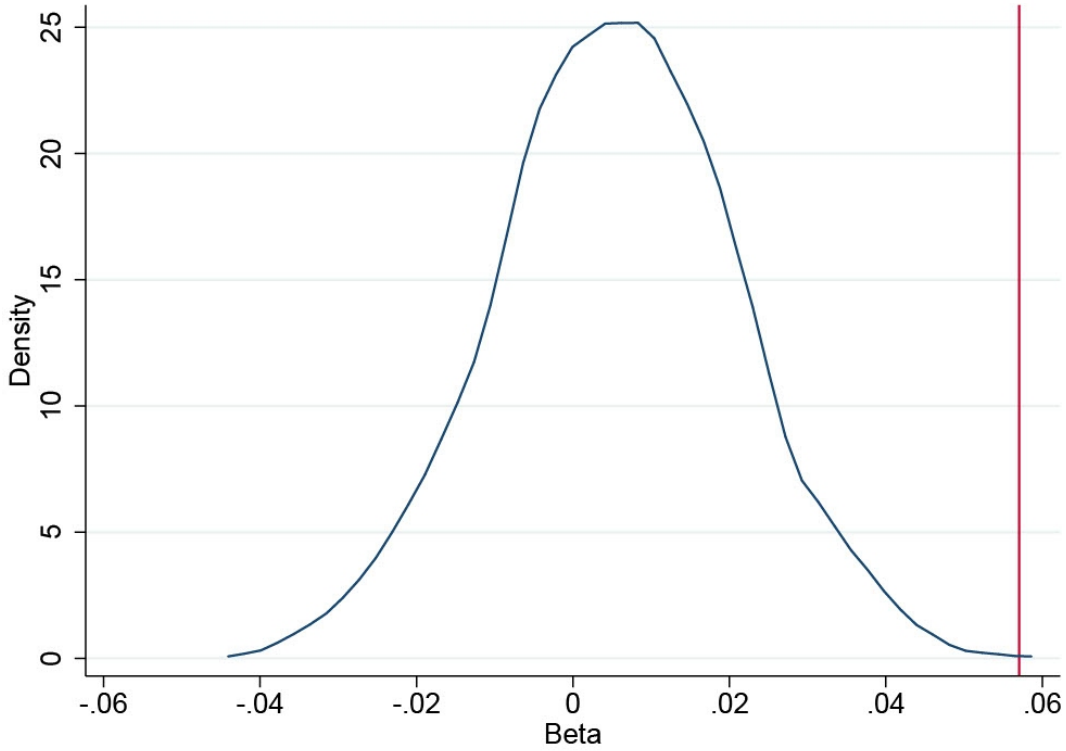
	(1)
	Treatment area: Outlier in LA
Interaction term	0.057
No Clustering	(0.014)***
Clustering on LSOA	**(0.017)***
Clustering on LSOA by Time	(0.018)**
Clustering on LA	(0.020)**
Clustering on LA by Time	(0.020)**
Clustering on PFA	(0.023)*
Clustering on PFA by Time	(0.022)*
Fixed Effect	LA
Time Trend	LSOA
Observations	3,742,962

*Notes: The table displays the estimate for the interaction term on the impact on burglaries estimated using Equation 1 with standard errors in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively.*

As alternatives we consider clustering at the local authority level or the police force area. Although the estimates lose some precision when larger clusters are considered (clustering at local authority (police force area) level involves 348 (42) clusters), our estimates do not fall below the 5% level of significance. Additionally, we also consider that the errors terms might be correlated over time (108 clusters) and cluster both at the geographical and time level. Again this is found to have little effect on the statistical inference which remains at the 5% level apart for PFA clusters.

In Figure 4, we additionally use randomisation inference as a falsification exercise whereby the indicator for a high South Asian neighbourhood is randomly allocated within a local authority. We then estimate the base model with this placebo indicator and replicate the analysis a 1,000 times. If some unknown characteristics of neighbourhoods was driving our results, we would expect to observe a substantial fraction of estimates to be larger than the one we report in our main analysis. However, as we can see in Figure 4 this is not the case with only 1 replication producing an estimate above the baseline value.

Figure 4: Distribution of placebo estimates, 1000 replications



Notes: The figure presents the empirical distribution of placebo estimates for the definition of treated neighbourhoods within local authorities. The cumulative distribution functions are based on 1,000 replications of Equation 1 with randomised treatment status of neighbourhoods. The vertical line corresponds to the value of the original estimate.

3.2 Changes in police behaviour

Police forces might be aware of the relationship between gold price and displacement of crime to Asian neighbourhoods, highlighted above, and may relocate police officers accordingly. While this does not invalidate our identification strategy, the estimates reported in our main analysis would then be under-estimates. While we do not have information on the deployment of police officer at the neighbourhood level we approximate their activity by the location of "stop and search" activities on a monthly basis. In particular, the Police and Criminal Evidence Act 1984 allows officers to stop and search on suspicion of carrying objects for use in theft or stolen goods. This is the second most popular legislation used to conduct searches (15% of all searches) and variations in this type of search could indicate that police officers focus changes, and thus affect displacement to other crimes.

Table 2: Impact of Gold Price on Stop and Search Activities

	Searches			
	All	Theft-related	All	Theft-related
High South Asian neighbourhood ×	-0.063	-0.006	-0.061	-0.005
Gold Price	(0.034)	(0.014)	(0.034)	(0.014)
High South Asian neighbourhood	0.712**	0.221*	0.000	0.000
	(0.251)	(0.106)	(.)	(.)
Gold Price	0.232***	-0.059***	0.231***	-0.059***
	(0.008)	(0.003)	(0.008)	(0.003)
Fixed Effect	LA	LA	LSOA	LSOA
Quadratic Time Trend	LSOA	LSOA	LSOA	LSOA
Observations	1,885,950			

*Notes: The table displays estimates of the impact on "stop and search" conducted by the police estimated using Eq. 1 of the main paper. Regressions also control for seasonality via monthly dummies. Standard errors adjusted for clustering at the LSOA level in parentheses. */**/*** denote statistical significance on the 10%, 5% and 1% level respectively.*

Using equation (1) in the main paper, we estimate whether variations in gold price relate to the (inverse hyperbolic sine) total number of "stops and searches" and to the "stops and searches" related to theft and stolen goods. Results are reported in Table 2. Within a local authority, high South Asian neighbourhoods appear to have a greater police presence - as approximated by total stop and search. Stops and searches appear to be positively correlated with variations in gold price, but this may not be related to redeployment of police officers to prevent burglaries since theft related stop and search are negatively correlated with gold price. Finally, neither for total search or theft-related search do we observe a change in search specifically in South Asian neighbourhoods. The point estimates are close to 0, and not statistically significant. Altogether, using stop and search to approximate police officer allocation and focus, we do not find any evidence that the police react to variations in gold price differentially in South Asian and other neighbourhoods. Our main estimates are thus likely to represent changes in the behaviour of criminals rather than changes in the allocation or behaviour of the police.

4 References

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. Measuring economic policy uncertainty. *The Quarterly Journal of Economics* 131, no. 4 (2016): 1593-1636.

Crow, Kevin (2006). SHP2DTA: Stata module to converts shape boundary files to Stata datasets, *Statistical Software Components S456718*, Boston College Department of Economics, revised 17 Jul 2015.

Kirchmaier, T., M. Langella, and A. Manning (2021). Commuting for crime. *CEP Discussion Paper 1747*.

ONS (2006) Model-Based Estimates of ILO Unemployment for LAD/UAs in Great Britain. Guide for Users, *Office of National Statistics*

Picard, Robert (2010). GEONEAR: Stata module to find nearest neighbors using geodetic distances, *Statistical Software Components S457146*, Boston College Department of Economics, revised 14 Sep 2019.

Pisati, Maurizio (2018). SPMAP: module to visualize spatial data.

Vandeviver, C. and W. Bernasco (2019). 'Location, location, location': Effects of neighborhood and house attributes on burglars' target selection. *Journal of Quantitative Criminology*, 1-43.