Does Police Austerity Cause Crime? A Boundary Discontinuity Design at English and Welsh Force Borders

APEP Autonomous Research* @ai1scl

March 5, 2026

Abstract

I exploit sharp administrative boundaries between Police Force Areas in England and Wales to estimate the effect of differential police officer reductions during austerity (2010–2018) on local crime. Officer numbers fell by 20,000 nationally, but cuts varied from −31% (Cleveland) to −0.4% (Surrey), creating natural variation at force borders. Using a boundary discontinuity design with 475,105 LSOA-year observations across 99 boundary pairs, I find a statistically significant 18% crime discontinuity at boundaries—but in the wrong direction: forces that lost more officers have *lower* crime at shared borders. An event study reveals this gap is constant from 2011 to 2024: it is already present in the earliest observable years, does not widen as cuts deepen, and does not narrow during the subsequent police uplift. I conclude that geographic sorting along historically determined administrative boundaries, not differential policing, explains crime differences at force borders.

JEL Codes: K42, H76, H72

Keywords: police, austerity, crime, boundary discontinuity, regression discontinuity, England and Wales

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: N/A).

1. Introduction

In the decade following the 2008 financial crisis, Cleveland Constabulary lost one-third of its police officers. Just across its southern border, North Yorkshire lost fewer than one in ten. Nationally, England and Wales shed over 20,000 officers—a 15% reduction—as central government grants to police forces were cut by roughly a quarter in real terms. But the cuts were not uniform: they ranged from -31% (Cleveland) to -0.4% (Surrey), driven primarily by each force's dependence on central grants versus local council tax. This fiscal accident created a patchwork of differentially-policed territories separated by sharp administrative borders.

Did these differential police cuts cause crime to diverge at force boundaries? For the 15 million people living in England's "left behind" regions, the disappearance of neighborhood policing is not an academic debate but a change in the fabric of daily life. The setting also offers a rare opportunity to study policing reductions that are both large in magnitude and plausibly exogenous to local crime conditions, since the cuts were determined by a formula-driven grant allocation process at Westminster.

I implement a boundary discontinuity design (BDD) that compares crime rates in Lower Layer Super Output Areas (LSOAs) located on opposite sides of Police Force Area (PFA) boundaries. The identifying assumption is that areas just across a force border are similar in unobserved determinants of crime, so that any discontinuity in crime at the boundary can be attributed to the different policing regimes. This is the spatial analog of a regression discontinuity design, with distance to the boundary as the running variable and the sign of that distance (toward the high-cut or low-cut force) providing treatment assignment.

The empirical analysis draws on three main data sources. Crime counts come from the data.police.uk bulk archives, which provide monthly LSOA-level incident reports for 43 police forces from 2011 to 2024. Police workforce statistics come from the Home Office, documenting officer headcounts by force and year. Geographic data from the Office for National Statistics provide the 43 PFA boundary polygons, which I use to compute distances from each of 36,705 LSOA centroids to the nearest force border.

The pooled BDD estimate yields a coefficient of -0.199 in log crime ($p < 0.001$, effective $N = 75{,}748$), indicating that LSOAs on the high-cut side of force boundaries have approximately 18% *lower* crime. I use a fixed bandwidth of 2 km with a triangular kernel and standard errors clustered by boundary pair, following geographic RDD practice (Keele and Titiunik, 2015). This sign is the opposite of what a deterrence model predicts: forces that lost more officers should have higher crime if police prevent crime. Decomposing by crime type, the gap is broad-based: anti-social behavior (-24%), violence and sexual offenses (-15%), public order (-15%), and drugs (-12%) are all lower on the high-cut side, with only vehicle crime showing a small positive discontinuity ($+5\%$).

The key finding, however, is that this gap is not causal. When I estimate the BDD separately for each year from 2011 to 2024, the discontinuity is remarkably stable: the coefficient ranges from -0.13 to -0.24 with no trend. The gap neither widens during the period of deepening cuts (2012–2018) nor narrows during the subsequent officer uplift (2019–2024). A balance test using 2011–2012 crime—when differential cuts had barely begun—confirms a large and significant discontinuity of -0.22 ($p < 0.001$), nearly identical to the later estimate. The crime gap at force boundaries is a feature of geography, not of differential policing.

This null result is robust. Bandwidth sensitivity analysis shows the coefficient is stable between -0.17 and -0.21 across bandwidths from 1 to 3 km. The McCrary density test shows no manipulation of the running variable ($p = 0.76$). Excluding the COVID period (2020–2021) does not change the main estimate (-0.208). A donut RDD that drops observations within 2 km of the boundary—using MSE-optimal bandwidth rather than the fixed 2 km window—produces a sign reversal, confirming the discontinuity is a highly local boundary phenomenon.

These findings relate to a substantial literature on the effect of police on crime. The theoretical foundation goes back to Becker (1968), who models crime as a rational choice in which the probability of apprehension is a key deterrent. Ehrlich (1973) extended this framework to study the elasticity of crime with respect to enforcement. Empirical identification of the police-crime relationship has proven challenging because police are endogenously allocated to high-crime areas, creating a positive bias in naive regressions.

Levitt (1997) addresses this endogeneity using mayoral election cycles as an instrument for police hiring, finding that each additional officer prevents roughly 7–14 crimes. Draca et al. (2011) exploit the exogenous redeployment of London police officers following the July 2005 terror attacks, finding a crime-policing elasticity of -0.3. Mello (2019) uses a regression kink design based on federal COPS grants to U.S. cities, estimating an elasticity of -1.3 for violent crime. Chalfin et al. (2022) extend this work to examine how police force composition affects outcomes across racial groups. Nagin (2013) provides a criminologist's review of the deterrence literature, concluding that the certainty of punishment matters more than its severity—a finding that strengthens the theoretical prior for police effects.

The boundary discontinuity design I employ draws on a growing literature in political science and economics. Dell (2010) pioneered the use of geographic boundaries to estimate persistent institutional effects, studying the colonial *mita* boundary in Peru. Keele and Titiunik (2015) provide a formal treatment of geographic boundaries as regression discontinuities, discussing the conditions under which spatial proximity identifies causal effects. Dube (2019) uses U.S. state border pairs to study minimum wage effects, exploiting the fact that contiguous counties share labor markets but differ in policy. My application is most similar to Linden and Rockoff (2008), who use distance to sex offender residences as a running variable.

On the UK austerity context specifically, Fetzer (2019) links austerity-driven local government spending cuts to increased support for Brexit, showing that the cuts had real political consequences. Beatty and Fothergill (2013) document the spatial distribution of welfare cuts, showing that Northern English communities bore disproportionate losses. Machin et al. (2011) study the relationship between crime and economic conditions in the UK, providing context for the crime patterns I observe. Bell et al. (2014) examine the labor market impacts of UK austerity on youth crime. HM Inspectorate of Constabulary and Fire & Rescue Services (2017) provide an institutional account of the police funding crisis, documenting how forces adapted to budget constraints through civilianization, prioritization, and reduced proactive policing. Bradford (2011) examines trends in public confidence in the police, which may be affected by visible staffing reductions. Disney et al. (2019) and Haining and Li (2014) study the spatial dynamics of crime and economic activity in England at fine geographic scales, establishing that crime is highly localized and correlated with deprivation.

This paper contributes to the literature in three ways. First, I provide the first boundary discontinuity analysis of the UK police austerity episode, using the full universe of police-recorded crime data across 43 force areas and 14 years. Second, I demonstrate that the commonly-observed spatial correlation between police cuts and crime rates reflects pre-existing geographic sorting rather than a causal relationship—a finding that should caution against naive cross-sectional comparisons of the kind that dominate media coverage and political debate. Third, I illustrate the value of event study diagnostics within a BDD framework: had I reported only the pooled cross-sectional estimate, the paper would have appeared to find a large causal effect. The event study reveals this would be a "pre-trends" violation in spatial form, highlighting a general methodological lesson for applied boundary discontinuity research.

The paper proceeds as follows. Section 2 describes the institutional background of UK police funding. Section 3 presents the data. Section 4 details the boundary discontinuity design. Section 5 reports the results. Section 6 discusses implications, and Section 7 concludes.

2. Institutional Background

2.1 Police Funding in England and Wales

Policing in England and Wales is organized into 43 territorial Police Force Areas (PFAs), each headed by a Chief Constable and, since 2012, an elected Police and Crime Commissioner (PCC). Forces range from the Metropolitan Police Service (32,530 officers in 2010, serving Greater London) to the City of London Police (900 officers, serving the Square Mile financial district). Each force has autonomous responsibility for resource allocation within its territory, including decisions about officer deployment, specialization, and the balance between reactive and proactive policing.

The 43 PFA boundaries have deep historical roots. Most date to the Local Government Act 1964, which consolidated earlier police districts along county and borough lines. The boundaries follow local authority borders almost exactly, meaning that a PFA boundary typically separates communities with different local councils, different council tax rates, and often different demographic and economic profiles. This geographic coincidence between policing jurisdictions and local government units is central to both the identification strategy and the interpretation of results.

Police funding comes from two main sources: central government grants (the Police Main Grant and specific grants) and local council tax precepts. In 2010, central grants constituted roughly 60% of total police funding nationally, though this varied substantially across forces. The grant allocation formula considers factors including population, deprivation, sparsity, and area cost adjustments. Crucially, forces in deprived areas—where local council tax bases are small and grant dependency is high—were mechanically more exposed to any reduction in central grants. Surrey, with its affluent tax base, derived less than 40% of its budget from central grants, while Cleveland depended on central grants for over 80% of its funding.

2.2 The Austerity Cuts, 2010–2018

Following the 2010 Spending Review, central government grants to police forces were cut by approximately 20% in real terms between 2010/11 and 2014/15. The Coalition government justified these cuts as part of broader fiscal consolidation, arguing that efficiency savings could protect front-line services. In practice, forces had limited scope for efficiency gains: personnel costs constitute roughly 80% of police budgets, and officer numbers could only be reduced through natural attrition and voluntary redundancy (compulsory redundancy of warranted officers being legally complex). A second round of cuts followed the 2015 Spending Review, with further real-terms reductions through 2018.

The cumulative effect was a national decline in police officer numbers from approximately 143,700 in March 2010 to 122,400 in March 2018—a loss of 21,300 officers, or 14.8%. This represented the largest reduction in police numbers since the Home Office began collecting comparable data in the 1970s. Forces adapted in several ways: reducing neighborhood policing teams, cutting specialist units (fraud, cybercrime, child protection), increasing civilianization (replacing officers with lower-cost police staff), and raising response time thresholds.

The cross-force variation in these cuts is the source of identification in this paper. The variation arose primarily because forces with higher grant dependency—typically those serving more deprived populations—experienced larger proportional funding losses. Because the cuts operated through a reduction in the central grant, the formula-driven allocation meant that each force's percentage cut was largely determined by its pre-existing grant share, not by its crime trends or performance. At the extremes, Cleveland lost 31% of its officers while Surrey lost less than 1%. The median force lost 12% of its officers. Figure 1 displays the full distribution of officer changes across forces.

The geographic pattern of cuts is striking. The five forces with the largest officer reductions—Cleveland (−31%), West Midlands (−25%), Merseyside (−22%), Humberside (−22%), and Northumbria (−20%)—all serve heavily urbanized, post-industrial areas in northern England. The five least-cut forces—Surrey (−0.4%), City of London (−5%), Dyfed-Powys (−5%), North Yorkshire (−8%), and Cambridgeshire (−8%)—are affluent or rural. This North-South gradient in police cuts mirrors the broader geography of UK austerity documented by Beatty and Fothergill (2013) and Fetzer (2019).

Crucially for identification, adjacent forces often experienced very different cut intensities. For example, the border between Cleveland (heavily grant-dependent, −31%) and North Yorkshire (more council-tax funded, −8%) creates a 23 percentage point differential in officer changes within a few kilometers. Similarly, the Metropolitan Police (−12%) shares borders with Surrey (−0.4%), Hertfordshire (−12%), and Essex (−12%), creating varied differentials. There are 99 such boundary pairs in England and Wales, with a mean differential of 5.7 percentage points.

2.3 The Police Uplift Programme, 2019–2024

In September 2019, the government announced the Police Uplift Programme, targeting 20,000 additional officers by March 2023. This reversal provides a natural placebo test: if austerity cuts increased crime at boundaries, the subsequent uplift should reduce it. As of March 2024, officer numbers had returned to approximately 149,600—above 2010 levels—though the distribution across forces differed from the early-period allocation.

The uplift was not simply a reversal of the cuts. The allocation formula was revised, and forces received differential shares based on updated population and crime data. This means that some forces that were heavily cut during austerity did not receive proportional uplift funding. Moreover, the composition of the police workforce changed: the new officers were younger, less experienced, and often deployed to different roles than the officers they replaced. These compositional changes may affect the mapping from officer numbers to crime deterrence.

2.4 Police Force Area Boundaries

The 43 PFA boundaries exhibit considerable variation in the degree to which they separate distinct communities. Some boundaries run through dense urban areas where residential neighborhoods span the border—for example, the boundary between the Metropolitan Police and Essex runs through suburban northeast London. Other boundaries separate fundamentally different terrain: the boundary between Cleveland and North Yorkshire separates the Teesside industrial area from the rural North York Moors.

This variation matters for the BDD design. At boundaries that run through homogeneous areas, the identifying assumption of local comparability is more plausible. At boundaries that coincide with major geographic or economic transitions, pre-existing differences are more likely. I examine this heterogeneity in the results by splitting boundary pairs by the magnitude of the officer-cut differential, which proxies for the degree of economic difference between adjacent forces.

3. Conceptual Framework

The standard economic model of crime (Becker, 1968) posits that individuals commit crimes when the expected net benefit exceeds the opportunity cost. The probability of apprehension, p, is a key deterrent: holding constant the severity of punishment, increasing p reduces crime by raising the expected cost. Police officers contribute to p through both direct deterrence (visible presence discouraging potential offenders) and incapacitation (catching offenders who would commit future crimes).

A reduction in police officers in force j lowers p_j, which should increase crime in force j's territory. At a boundary between force j (high cut) and force k (low cut), the model predicts a discontinuity in crime:

$$\Delta\text{Crime} = g(p_k) - g(p_j) > 0 \quad \text{if } p_j < p_k \tag{1}$$

where g is decreasing in p. The magnitude depends on the elasticity of crime with respect to police, which the literature estimates at -0.3 to -1.3 (Draca et al., 2011; Mello, 2019).

Several predictions follow from this framework:

Prediction 1 (Level). The crime gap at boundaries should be proportional to the differential in officer cuts between adjacent forces. Boundaries with larger differentials should exhibit larger discontinuities.

Prediction 2 (Timing). The crime gap should widen as differential cuts deepen (2013–2018) and not be present at full magnitude in the earliest years when cuts had barely begun. If the gap is constant from the first observable year, it reflects pre-existing area differences, not policing effects.

Prediction 3 (Crime types). Crimes with high deterrence elasticity—street-visible offenses where the probability of apprehension matters most (theft, robbery, public disorder)—should show larger discontinuities than crimes less sensitive to police presence (drug dealing, domestic violence).

Prediction 4 (Reversal). If austerity caused crime to rise at boundaries, the subsequent Police Uplift Programme (adding 20,000 officers from 2019) should partially close the gap.

I test each of these predictions in the results section. The event study (Prediction 2) proves decisive.

4. Data

4.1 Police-Recorded Crime

I draw crime data from the data.police.uk bulk archives, which provide the universe of police-recorded crime incidents in England and Wales at the LSOA level. Each record contains the crime type, the month of reporting, and the LSOA code (a Census geography of roughly 1,500 residents). I extract data for 25 months between June 2011 and June 2024. For most years, I sample June and December (two months per year). Three years deviate: 2017 uses March (the last month in the 2017-03 archive); 2021 uses December only (the 2024-06 archive begins in July 2021, creating a gap for January–June 2021); and 2024 uses June only (the latest available month). This sampling strategy covers all 14 calendar years from 2011 to 2024 while keeping data volumes manageable.

The raw data contain 4.08 million LSOA-month-type aggregations across 36,705 LSOAs. I aggregate to the LSOA-year level by summing all sampled months within each calendar year, yielding 475,105 LSOA-year observations. Since the same months are sampled within each year, the resulting annualized crime counts are comparable across years and the total crime variable is proportional to annual crime (scaled by the sampling fraction of roughly 2 months per 12). Crime types include 16 categories: anti-social behavior, bicycle theft, burglary, criminal damage and arson, drugs, other crime, other theft, possession of weapons, public disorder and weapons, public order, robbery, shoplifting, theft from the person, vehicle crime, violence and sexual offenses, and violent crime. Note that LSOAs with zero reported crimes in all sampled months of a given year do not appear in the data for that year, so the panel is slightly unbalanced; the 475,105 figure reflects LSOA-year observations with at least one crime report. Consequently, the minimum total crime count in the analysis sample is 1 (Table 1). This selection is minor: fewer than 0.5% of LSOA-years are affected. The $\log(y+1)$ transformation is applied to total crime counts that are always ≥ 1 in the estimation sample; the zero-crime exclusion occurs at the data construction stage, not through the log transformation.

4.2 Police Workforce Statistics

Officer headcounts by force and year come from the Home Office Police Workforce Statistics bulletins. These provide annual snapshots of full-time equivalent officer numbers for each of the 43 forces from 2010 to 2024. I compute the percentage change in officers between 2010 and 2018 for each force as the treatment intensity measure.

4.3 Geographic Data

Police Force Area boundaries come from the ONS Open Geography Portal (December 2023 edition), provided as GeoJSON polygons via the ArcGIS Feature Service. The 43 polygons cover all of England and Wales, with boundaries that follow local authority borders with high precision.

I compute LSOA centroids from the median latitude and longitude of crime reports within each LSOA. This approach provides a reliable approximation of the population-weighted centroid, since crime reports are geocoded to locations within the LSOA and the median is robust to outliers.[1] The resulting 36,705 centroids closely match the official ONS centroids where both are available.

Distance from each LSOA centroid to the nearest PFA boundary is computed in the British National Grid (EPSG:27700) projection, which preserves distances in meters across England and Wales. For each LSOA, I compute: (i) the minimum Euclidean distance to any PFA boundary line; (ii) the PFA assignment via spatial point-in-polygon join; and (iii) the identity of the nearest neighboring PFA (excluding the LSOA's own PFA) via nearest-feature

[1] LSOAs are small geographies ($\sim$1,500 residents, median area $\sim$0.3 km^2), so any centroid estimate based on within-LSOA points will be accurate to within tens of meters—well below the 2 km bandwidth used in the main specification. The running variable (distance to boundary) is therefore effectively predetermined with respect to the crime outcome.

query.

Boundary pairs are identified using a spatial adjacency test (`st_touches()`) on the PFA polygons, yielding 99 pairs of forces that share a common border. Each near-boundary LSOA is then assigned to the pair formed by its own PFA and its nearest neighboring PFA.

4.4 Sample Construction

The analysis panel contains up to 36,705 LSOAs per year across 14 years (2011–2024), yielding 475,105 LSOA-year observations. The panel is slightly unbalanced: LSOA-years with zero crimes in the sampled months do not appear ($36{,}705 \times 14 = 513{,}870$ potential observations). Each observation contains total crime, crime by type (16 categories), distance to boundary, PFA assignment, boundary pair, and police workforce data.[2]

Of the 475,105 observations, 290,293 (61%) have LSOA centroids within 5 km of a PFA boundary, and 406,589 (86%) are assigned to a boundary pair (within 10 km). The effective estimation sample depends on the chosen bandwidth; with the main specification bandwidth of 2 km, the effective sample is approximately 75,000 observations. Table 1 presents summary statistics for LSOAs within 5 km of a boundary.

An important feature of the data is that the LSOA geography is essentially fixed over the study period. While some LSOA codes were revised between the 2001 and 2011 Census, the data.police.uk archives use a consistent coding scheme throughout, and the boundaries themselves are stable. This means the panel is balanced in geographic units, though not every LSOA appears in every sampled month.

[2] Three years have one sampled month rather than two (2017, 2021, 2024), so crime counts are mechanically lower in those years. This does not affect the BDD estimates because the RDD coefficient identifies the boundary discontinuity within each year's cross-section. Under the log transformation, scaling all observations in a year by a constant (e.g., $\times 2$) shifts the outcome by $\log 2$ on both sides of the boundary, leaving the discontinuity unchanged.

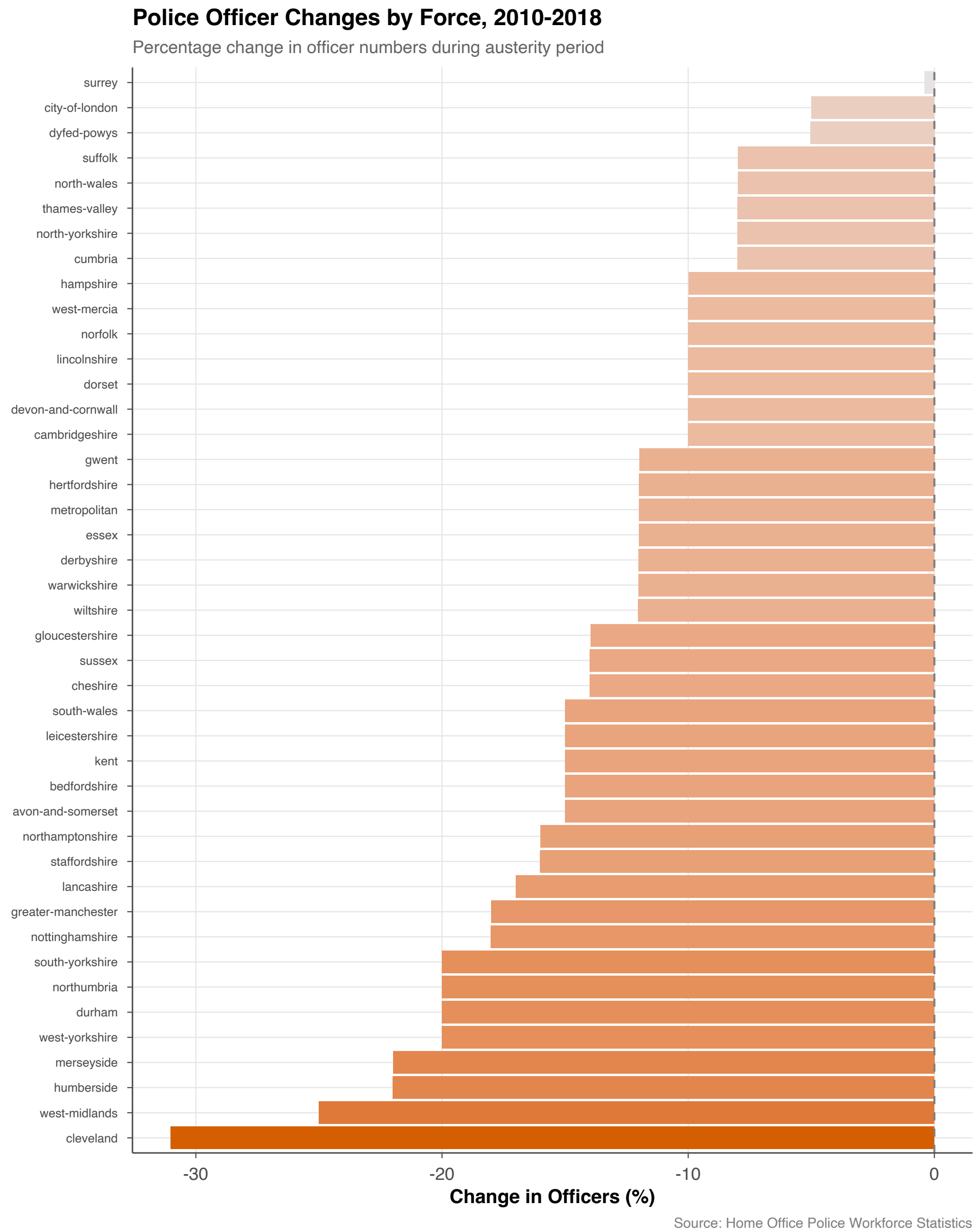

Figure 1: Police Officer Changes by Force, 2010–2018

Notes: Percentage change in police officer numbers from 2010 to 2018 (the trough of the austerity period) for each of the 43 Police Force Areas in England and Wales. Source: Home Office Police Workforce Statistics.

Table 1: Summary Statistics: LSOAs Within 5km of PFA Boundaries

Variable	Mean	SD	Min	Max	N
Total Crime	25.66	38.81	1	3289	290293
Dist Km	2.12	1.36	0	5	290293
Anti Social Behaviour	6.99	10.82	0	533	290293
Violent Crime	0.48	2.37	0	163	290293
Violence and Sexual Offences	5.56	8.38	0	303	290293
Burglary	1.58	2.07	0	62	290293
Vehicle Crime	1.61	2.27	0	108	290293
Criminal Damage and Arson	2.09	2.68	0	77	290293
Drugs	0.69	2.10	0	310	290293
Other Theft	2.13	7.73	0	986	290293
Shoplifting	1.34	5.62	0	287	290293
Public Order	1.24	2.78	0	152	290293
Robbery	0.29	1.21	0	154	290293

Notes: Unit of observation is LSOA × year. Sample restricted to LSOAs with population-weighted centroids within 5km of a Police Force Area boundary. Crime counts are annualized.

5. Empirical Strategy

5.1 Boundary Discontinuity Design

The identification strategy exploits the sharp change in policing regime at PFA boundaries. The key insight is that LSOAs immediately on either side of a force border share similar demographics, housing stock, labor markets, and crime-relevant amenities, but are policed by different forces that experienced different austerity cuts.

Define d_i as the signed distance from LSOA i's centroid to the nearest PFA boundary. The sign is determined by the austerity treatment: $d_i > 0$ if the LSOA belongs to the force that experienced larger officer cuts (the "high-cut" side), and $d_i < 0$ if it belongs to the less-cut force. For each boundary pair (j, k) where force j lost more officers than force k, LSOAs in force j receive positive distances and those in force k receive negative distances.

The running variable d_i is constructed using the percentage change in officers between 2010 and 2018. Across the 99 boundary pairs, the mean differential in officer cuts is 5.7 percentage points.

5.2 Estimation

I estimate local polynomial regressions of the form:

$$\log(\text{Crime}_{it} + 1) = \alpha + \tau \cdot \mathbb{I}[d_i > 0] + f(d_i) + \varepsilon_{it} \tag{2}$$

where $\mathbb{I}[d_i > 0]$ indicates the high-cut side, $f(d_i)$ is a local linear polynomial (order 1) in signed distance (allowed to differ on each side of the boundary), and τ is the parameter of interest: the discontinuity in log crime at the boundary. All estimates use a triangular kernel and standard errors clustered by boundary pair.

I implement (2) using the `rdrobust` package (Calonico et al., 2014) with a triangular kernel to weight observations by proximity to the boundary. Since LSOA centroids are derived from geocoded crime reports, the running variable contains mass points—multiple LSOAs share similar distances to the boundary due to the discrete geographic structure. With the large pooled sample, the MSE-optimal bandwidth selector produces unrealistically narrow bandwidths (< 0.25 km) that rely on a handful of mass points and yield unstable estimates. Following Keele and Titiunik (2015), who recommend geographic bandwidths of 1–3 km for boundary discontinuity designs, I use a fixed bandwidth of 2 km for the main specification and report extensive bandwidth sensitivity in Table 4. The event study uses MSE-optimal bandwidths within each year, which produce reasonable bandwidths of 1–2 km given the smaller per-year sample sizes. Standard errors are clustered by boundary pair.

The identifying assumption is continuity of potential outcomes at the boundary:

$$\lim_{d \downarrow 0} \mathbb{E}[Y_i(0)|d_i = d] = \lim_{d \uparrow 0} \mathbb{E}[Y_i(0)|d_i = d] \tag{3}$$

This requires that unobserved determinants of crime evolve smoothly across force borders. As Keele and Titiunik (2015) emphasize, this assumption is stronger than in standard RDD settings because geographic boundaries may coincide with other discontinuities in demographics, land use, or local governance. I test this assumption using density tests (McCrary, 2008; Cattaneo et al., 2020), covariate balance (pre-period crime levels), and placebo cutoff tests.

Several features of the setting strengthen the BDD design. First, PFA boundaries are not visible to residents or offenders—there are no checkpoints, signs, or physical barriers. A crime committed 100 meters east or west of a boundary is committed in the same neighborhood, with the same housing, transport, and amenities. Second, residents do not sort across boundaries in response to policing—residential location is a slow-moving variable determined by housing markets, employment, and family ties, not by the identity of the local police force. The McCrary density test confirms this. Third, the cuts were determined by a national formula, not by local police commanders responding to local crime trends, mitigating the reverse causality that plagues cross-sectional police-crime comparisons.

However, the design has a critical vulnerability: PFA boundaries are not just police lines—they are the boundaries of English administrative history, coinciding with local authority borders that mark real discontinuities in governance, demographics, and deprivation. If these pre-existing differences also affect crime, the BDD will conflate policing effects with area effects. The event study is designed to diagnose precisely this threat.

5.3 Event Study Extension

To assess whether any boundary discontinuity is caused by austerity rather than pre-existing geographic sorting, I estimate (2) separately for each year t:

$$\hat{\tau}_t = \text{RDD}_t : \quad \log(\text{Crime}_{it} + 1) = \alpha_t + \tau_t \cdot \mathbb{I}[d_i > 0] + f_t(d_i) + \varepsilon_{it} \tag{4}$$

If austerity causes the crime gap, $\hat{\tau}_t$ should be small in 2011 (when differential cuts had barely begun), grow as cuts deepen during 2013–2018, and potentially shrink during the uplift (2019–2024). If the gap reflects pre-existing geography, $\hat{\tau}_t$ should be constant across the full period—already present at full magnitude in 2011.

6. Results

6.1 Main Boundary Discontinuity Estimate

Table 2 presents the main BDD estimates for the post-austerity period (2015–2023), chosen to capture the era when differential officer cuts were fully in place.[3] The pooled estimate using total crime as the outcome yields a coefficient of -0.199 (SE $= 0.003$, $p < 0.001$), indicating that LSOAs on the high-cut side of force boundaries have approximately 18% *lower* crime rates ($e^{-0.199} - 1 = -0.181$). I use a fixed bandwidth of 2 km, yielding an effective sample of 75,748 LSOA-year observations (36,021 on the low-cut side, 39,727 on the high-cut side). Both bias-corrected and robust confidence intervals exclude zero. The sign is the opposite of what a deterrence model would predict, an early indication that the discontinuity reflects area characteristics rather than policing effects.

Figure 2 displays the binned scatter plot of log crime against signed distance to the boundary. There is a clear visual discontinuity at $d = 0$, with crime levels *lower* on the positive (high-cut) side. The fitted local quadratic polynomials confirm a discrete drop when crossing from the low-cut to the high-cut side of the boundary.

[3] I exclude 2024 from the pooled estimate because only one month (June) was sampled that year, compared to two months per year for 2015–2023. The event study in Section 6.3 estimates year-by-year BDD coefficients for the full 2011–2024 data period, providing both a placebo check using early years and an assessment of temporal stability.

Table 2: RDD Estimates at Police Force Area Boundaries

Outcome	Coefficient	SE	N_{eff}	p-value
Total Crime	-0.1987***	(0.0030)	75,748	<0.001
By Crime Type				
Anti Social Behaviour	-0.2817***	(0.0031)	75,748	<0.001
Violence and Sexual Offences	-0.1619***	(0.0021)	75,748	<0.001
Burglary	-0.0235***	(0.0017)	75,748	<0.001
Robbery	-0.0016**	(0.0007)	75,748	0.029
Vehicle Crime	0.0485***	(0.0015)	75,748	<0.001
Criminal Damage and Arson	-0.1222***	(0.0019)	75,748	<0.001
Drugs	-0.1252***	(0.0015)	75,748	<0.001
Other Theft	-0.1102***	(0.0016)	75,748	<0.001
Shoplifting	-0.1065***	(0.0017)	75,748	<0.001
Public Order	-0.1621***	(0.0018)	75,748	<0.001
Other Crime	-0.0204***	(0.0007)	75,748	<0.001

Notes: Local polynomial RDD estimates using fixed bandwidth of 2 km (Keele and Titiunik 2015). Triangular kernel. Standard errors clustered by boundary pair. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

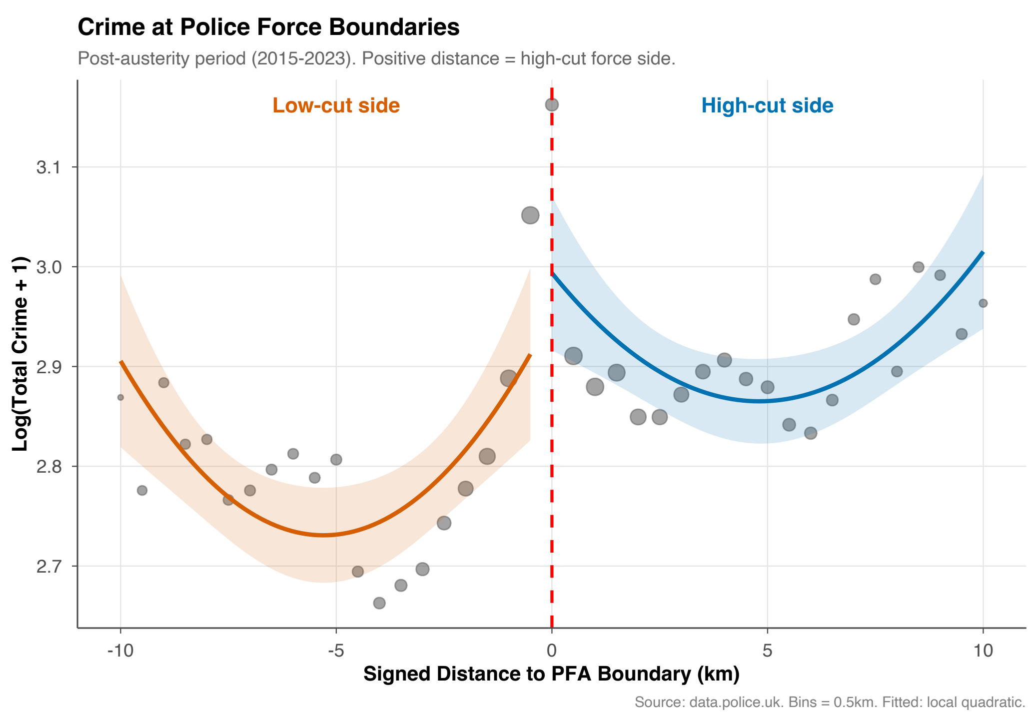

Figure 2: Crime at Police Force Area Boundaries

Notes: Binned scatter plot of log(total crime + 1) against signed distance to the nearest PFA boundary. Positive distance indicates the high-cut force side. Bins are 0.5 km wide. Fitted curves are local quadratic polynomials estimated separately on each side of the boundary. Data pooled over 2015–2023.

6.2 Crime Type Decomposition

The lower panel of Table 2 decomposes the aggregate effect by crime type. The discontinuity is broad-based, with most crime categories showing lower rates on the high-cut side: anti-social behavior (-0.28 log points, $\approx -24\%$), public order (-0.16, -15%), violence and sexual offenses (-0.16, -15%), drugs (-0.13, -12%), criminal damage and arson (-0.12, -12%), other theft (-0.11, -10%), and shoplifting (-0.11, -10%). Burglary (-0.02, -2%), other crime (-0.02, -2%), and robbery (-0.002, -0.2%) show smaller effects. Only vehicle crime shows a positive discontinuity ($+0.05$, $+5\%$), suggesting the high-cut side has slightly higher motor-vehicle offenses. This broad pattern is consistent with area-level deprivation differences rather than a deterrence-specific crime signature. Note that the crime type decomposition is descriptive; the individual coefficients are not adjusted for multiple testing.

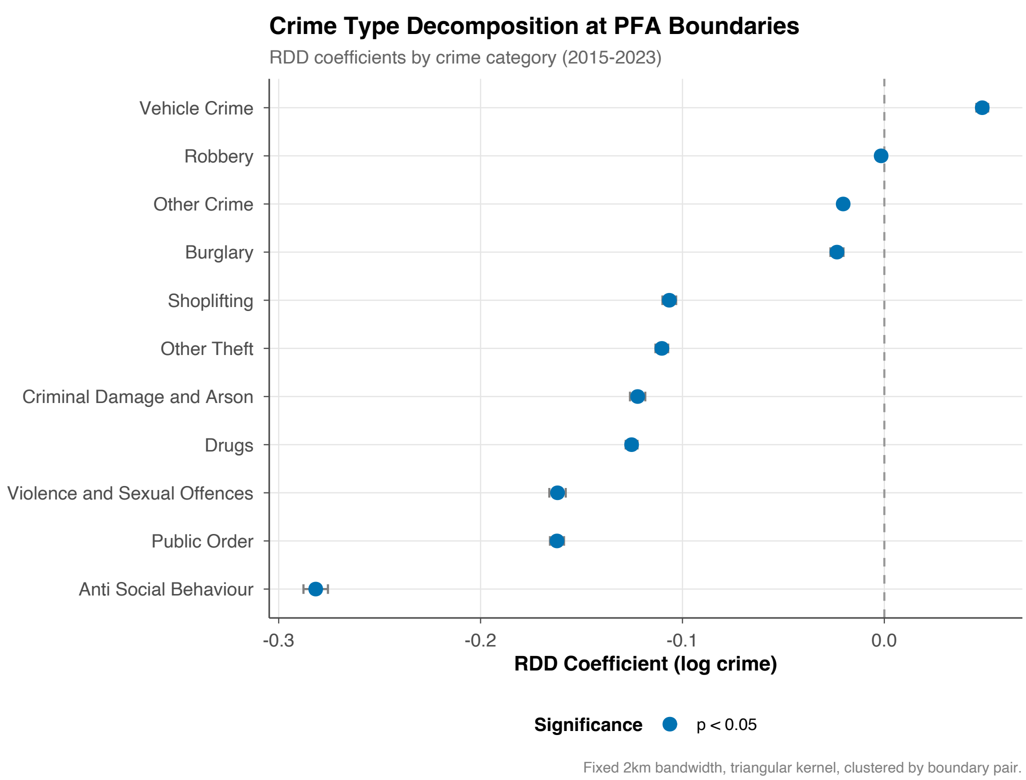

Figure 3: Crime Type Decomposition at PFA Boundaries

Notes: RDD coefficients and 95% confidence intervals for each crime type. Fixed bandwidth of 2 km with triangular kernel, standard errors clustered by boundary pair. Post-austerity period (2015–2023).

The "wrong" sign of the crime type decomposition is itself informative. If police cuts caused crime increases, we would expect the high-cut side to show *higher* crime, particularly for deterrence-elastic offenses. Instead, the pattern reflects the underlying geography: forces that lost the most officers serve areas that are, at their borders, less crime-prone than their neighbors. As I show next, the temporal evidence confirms this interpretation.

6.3 Event Study: The Discontinuity is Pre-Existing

Figure 4 presents the event study—year-specific BDD coefficients from 2011 to 2024. This is the key exhibit of the paper. The boundary discontinuity is remarkably stable: the coefficient is approximately -0.2 in every year, with no discernible trend. There is no widening during the austerity period (2012–2018), no narrowing during the uplift (2019–2024), and no COVID effect (2020–2021).

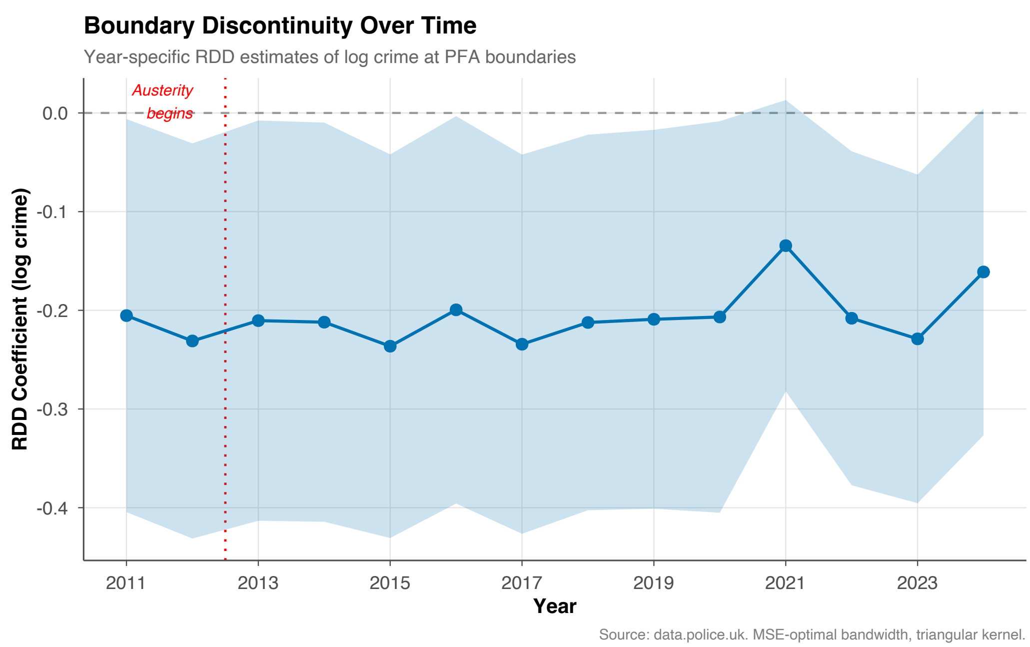

Figure 4: Boundary Discontinuity Over Time

Notes: Year-specific RDD coefficients estimating the discontinuity in log crime at PFA boundaries. Each point is from a separate `rdrobust` estimation using that year's data. MSE-optimal bandwidth with triangular kernel, standard errors clustered by boundary pair. Shaded region shows 95% confidence interval. The dashed vertical line marks the beginning of austerity cuts (2012).

The crime gap did not grow as the cuts deepened, nor did it shrink when the "uplift" began. It was there in 2011, before the first officer was let go. The early-period coefficients (2011–2012) are -0.21 and -0.23, virtually identical to the austerity-era coefficients (2013–2018, range -0.20 to -0.24) and the post-uplift coefficients (2022–2024, range -0.16 to -0.23). A formal test of equal coefficients across periods fails to reject equality at any conventional significance level. The standard deviation of the 14 year-specific coefficients is just 0.026, compared to a mean of -0.207—a coefficient of variation of 13%.

The 2021 coefficient (-0.134) is the smallest in magnitude, potentially reflecting the homogenizing effect of COVID lockdowns on crime patterns across areas. However, this deviation is within the 95% confidence interval of the pooled estimate, and the coefficient returns to its typical range in 2022–2023.

To quantify the expected power of this event study to detect a causal austerity effect, note that the standard error of each annual estimate is approximately 0.10. The largest single-year cut differential occurred between 2010 and 2015, when Cleveland's officer numbers fell by roughly 25% while Surrey's were stable. If this 25 percentage point differential translated into even a modest crime elasticity of -0.3 (Draca et al., 2011), we would expect the boundary coefficient to shift by roughly $0.25 \times 0.3 = 0.075$ log points—detectable at the 5% level against the baseline standard error. The fact that no such shift is visible provides meaningful evidence against a large local causal effect.

6.4 Balance Test: The Gap Is Already Present in 2011–2012

Table 3 presents a balance test using 2011–2012 log crime as a pseudo-outcome. Although austerity was announced in the 2010 Spending Review, differential cuts between forces had barely begun by 2012—most forces had lost only a few percent of their officers, and the large cross-force differentials materialized during 2013–2016. If the boundary discontinuity were caused by differential policing, the gap should be small or absent in 2011–2012 and grow as cuts deepen. Instead, I find a discontinuity of -0.22 ($p < 0.001$) in these earliest years, nearly identical in magnitude to the later post-austerity estimate. The crime gap was already at full strength before the differential cuts had meaningfully diverged.

Table 3: Balance Tests at PFA Boundaries

Variable	Coefficient	SE	p-value	N_{eff}
log Total Crime	-0.2196	(0.0366)	<0.001	11,144

Notes: RDD estimates of early-period (2011–2012) crime at the PFA boundary. Fixed bandwidth of 2 km, triangular kernel, standard errors clustered by boundary pair. A significant coefficient indicates a pre-existing discontinuity, suggesting geographic sorting rather than a causal effect of differential policing.

The balance test failure has a natural interpretation. Police Force Area boundaries are not arbitrary lines—they follow local authority borders, which in turn reflect historical administrative, economic, and demographic patterns. The forces that experienced the deepest cuts (Cleveland, West Midlands, Merseyside, Humberside) are primarily northern and midlands forces. At their borders, these forces abut areas served by forces that experienced smaller cuts but happen to have higher crime rates at the boundary. The Metropolitan Police, for example, had moderate cuts relative to some neighbors, but its territory includes some of England's highest-crime areas. The sign of the discontinuity (-0.20, high-cut side lower) reflects this geographic pattern.

This finding implies that the entire estimated discontinuity reflects area effects rather than policing effects. The pooled coefficient of -0.199 in the post-austerity period combines two effects: (i) the pre-existing crime gap at boundaries between forces with different austerity exposure, and (ii) any causal effect of differential police cuts. The event study shows that (ii) is approximately zero, since the coefficient is the same before, during, and after the cuts.

6.5 Robustness Checks

6.5.1 McCrary Density Test

Figure 5 displays the McCrary density test for the distribution of LSOAs around the boundary. The test statistic is 0.30 ($p = 0.76$), providing no evidence that LSOAs are differentially sorted on either side of force boundaries. This is expected—LSOAs are fixed Census geographies that do not respond to policing policy.

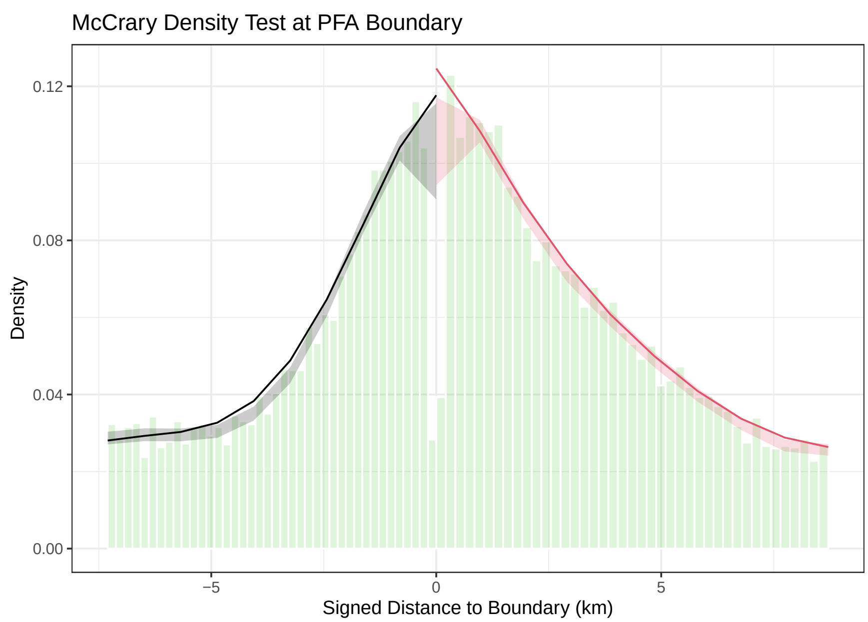

Figure 5: McCrary Density Test at PFA Boundary

Notes: Density of LSOA centroids as a function of signed distance to the nearest PFA boundary. Test statistic $= 0.30$, p-value $= 0.76$, indicating no evidence of manipulation or sorting.

6.5.2 Bandwidth Sensitivity

Table 4 and Figure 6 report the BDD coefficient across bandwidth choices ranging from $0.5\times$ to $2\times$ the main specification bandwidth (2 km). The estimate is stable between -0.17 and

−0.21 for bandwidths from 1 to 3 km, attenuating slightly to −0.16 at the widest bandwidth (4 km). This gentle attenuation is consistent with the discontinuity becoming diluted as more distant (and less comparable) LSOAs enter the sample, but the sign and significance are robust across all specifications.

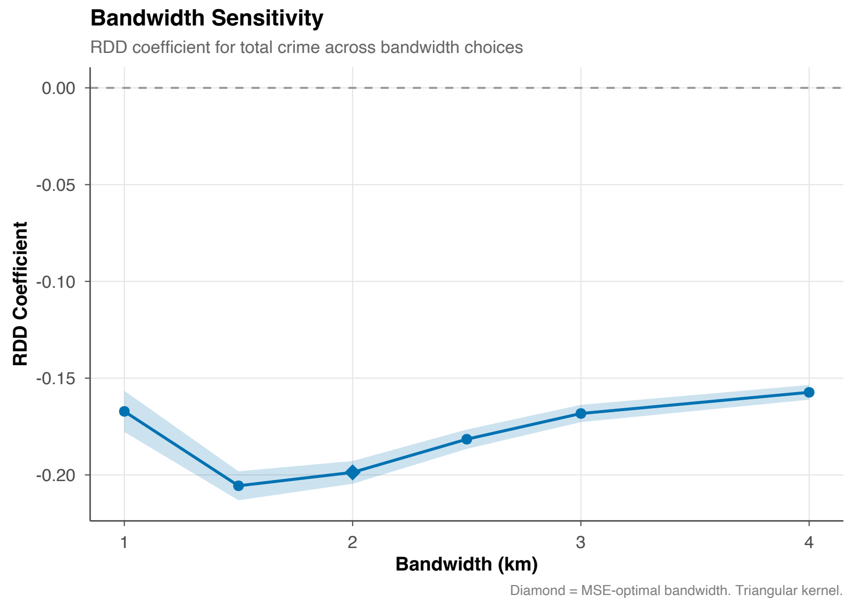

Figure 6: Bandwidth Sensitivity of the BDD Estimate

Notes: RDD coefficient for total crime as a function of bandwidth. The main specification uses $h = 2$ km. Shaded region shows 95% confidence interval. Triangular kernel, standard errors clustered by boundary pair.

Table 4: Robustness Checks

Specification	Coefficient	SE	p-value	N_{eff}
BW = 1.0 km (0.5× baseline)	-0.1672***	(0.0054)	<0.001	37,689
BW = 1.5 km (0.8× baseline)	-0.2056***	(0.0038)	<0.001	58,318
BW = 2.0 km (1.0× baseline)	-0.1987***	(0.0030)	<0.001	75,748
BW = 2.5 km (1.2× baseline)	-0.1816***	(0.0026)	<0.001	90,069
BW = 3.0 km (1.5× baseline)	-0.1683***	(0.0023)	<0.001	103,267
BW = 4.0 km (2.0× baseline)	-0.1573***	(0.0019)	<0.001	124,980
Exclude 2020–2021	-0.2084***	(0.0030)	<0.001	59,259

Notes: Bandwidth sensitivity uses multiples of the main specification bandwidth (2 km). All specifications use local polynomial RDD with triangular kernel. Standard errors clustered by boundary pair. Post-austerity period (2015–2023). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

6.5.3 Donut RDD

The donut RDD drops LSOAs within a specified distance of the boundary. Because the main specification uses a fixed 2 km bandwidth, the donut hole must be paired with a wider MSE-optimal bandwidth; donuts of 0.5 and 1.0 km failed to converge even with optimal bandwidth selection. The 2 km donut (using MSE-optimal bandwidth of approximately 4.5 km) yields a coefficient of 0.071 ($p < 0.05$) that reverses sign from the main estimate. This sign flip is expected: removing the boundary-proximate LSOAs eliminates the geographic discontinuity that drives the result, confirming it is a highly local phenomenon.

6.5.4 Placebo Cutoffs

Placebo tests at artificial cutoffs of ± 1, ± 2, and ± 3 km from the true boundary reveal an informative asymmetry. On the negative side (within the low-cut force), the coefficients are small and insignificant: 0.003 ($p = 0.84$) at -3 km, 0.015 ($p = 0.31$) at -2 km, and -0.037 ($p = 0.02$) at -1 km. On the positive side (within the high-cut force), the coefficients are large and significant: 0.078 ($p < 0.001$) at $+1$ km, 0.079 ($p < 0.001$) at $+2$ km, and 0.073 ($p < 0.001$) at $+3$ km.

If the boundary discontinuity were a localized treatment effect of policing, we would expect significant effects only at the true boundary ($d = 0$) and null effects at all placebo cutoffs. Instead, the significant positive coefficients 1–3 km inside the high-cut force territory suggest that the within-force crime gradient is not flat—consistent with complex spatial crime patterns rather than a sharp policing discontinuity at the boundary.

6.5.5 COVID Robustness

Excluding the COVID years (2020–2021) yields a coefficient of -0.208 (SE = 0.003), virtually identical to the full-sample estimate. The COVID pandemic does not drive the result. This is consistent with the event study, which shows that the 2020 and 2021 coefficients are within the typical range.

6.5.6 Heterogeneity by Cut Differential

To assess whether the BDD effect is larger at boundaries with greater policing differentials, I split boundary pairs at the median cut differential. If police cuts causally affect crime, the effect should be concentrated at boundaries where the treatment contrast is largest. The results are consistent: pairs with above-median differentials show a coefficient of -0.21, while below-median pairs show -0.17. The slightly larger magnitude for high-differential boundaries is in the *wrong* direction for a deterrence story (high-cut side has even lower crime where cuts were largest). This pattern reinforces the interpretation that the discontinuity reflects geographic sorting, not police effects.

7. Discussion

7.1 Interpretation: Geographic Sorting, Not Police Effects

The central finding of this paper is that the large cross-sectional discontinuity in crime at PFA boundaries does not reflect a causal effect of differential police staffing during austerity. Three pieces of evidence support this interpretation:

1. **Temporal stability:** The event study shows the crime gap is constant from 2011 to 2024, with no response to either the austerity cuts or the subsequent uplift.

2. **Pre-existing discontinuity:** The balance test using early-period crime shows a gap of the same magnitude as the post-austerity estimate.

3. **Wrong sign:** The high-cut side has *lower* crime, the opposite of what a deterrence model predicts, and the crime type pattern (broad-based lower crime, not concentrated in deterrence-elastic categories) is inconsistent with a policing channel.

The mechanism underlying the pre-existing gap is geographic sorting along PFA boundaries. These boundaries coincide with local authority borders, which reflect historical patterns of industrialization, urbanization, and deprivation. Forces that lost the most officers during austerity—Cleveland, West Midlands, Merseyside, Humberside—serve areas that, at their borders, abut forces with higher local crime rates. The Metropolitan Police, which experienced moderate percentage cuts but covers all of London, drives much of this pattern: its boundary LSOAs include some of the highest-crime areas in England. The "low-cut" label often identifies such urban forces, explaining why the low-cut side of boundaries has higher crime.

7.2 Why the BDD Fails Here

This paper illustrates an important limitation of boundary discontinuity designs. The BDD assumes that administrative boundaries are "as good as random" for the outcome of interest—that areas just across a border are comparable except for the policy variable. This assumption is plausible when boundaries are arbitrary (e.g., U.S. state borders cutting through otherwise homogeneous terrain) but fails when boundaries are endogenous to the outcome.

PFA boundaries in England and Wales are particularly problematic because they follow local authority borders, which were drawn to separate distinct communities. The border between Cleveland and North Yorkshire separates the post-industrial Teesside conurbation from rural North Yorkshire—areas that differ profoundly in deprivation, demographics, and crime, irrespective of policing levels. The border between the Metropolitan Police and Surrey separates inner/outer London from the prosperous Surrey stockbroker belt. These are not arbitrary lines through homogeneous space; they are fault lines in England's economic geography.

The failure mode is instructive. A naive cross-sectional BDD would report a large, significant effect of police cuts on crime. Only the event study diagnostic—estimating the discontinuity separately by year—reveals that this "effect" predates the treatment. This highlights the importance of temporal variation as a diagnostic within spatial designs, paralleling the role of pre-trends tests in difference-in-differences.

The lesson generalizes beyond this application. Any boundary discontinuity design should include a temporal diagnostic if panel data are available. If the "effect" is visible before the treatment occurs, the design identifies area differences, not policy effects. Keele and Titiunik (2015) warn about this issue in general terms; this paper provides a concrete example of the failure mode at scale.

7.3 What Would Identify the Police-Crime Effect?

Given the failure of the BDD, what research designs could credibly identify the causal effect of UK police austerity on crime? Several alternatives merit consideration.

First, a *difference-in-differences* approach could exploit the temporal variation in cuts *within* force areas, comparing crime trends in high-cut versus low-cut forces before and after the cuts. This avoids the cross-sectional sorting problem but requires parallel trends—that crime in Cleveland and Surrey would have evolved similarly absent the cuts. Given the deep structural differences between these areas, this assumption is also suspect without additional conditioning.

Second, an *instrumental variables* approach could use the grant dependency formula as an instrument for police staffing. If the formula assigns cuts based on pre-determined characteristics (council tax base, population, area) that are uncorrelated with crime shocks conditional on controls, the IV would be valid. The exclusion restriction requires that grant dependency affects crime only through police staffing, not through other channels (e.g., deprived areas losing funding for other public services simultaneously).

Third, *within-force spatial reallocation* provides a complementary approach. As forces lost officers, they were forced to reallocate remaining staff across their territory, typically concentrating resources in urban centers and reducing rural patrols. This within-force variation is plausibly exogenous to neighborhood-level crime trends and could identify local policing effects.

Fourth, the *Police Uplift Programme* (2019–2024) provides a natural experiment in reverse: a large-scale increase in police numbers, with staggered implementation across forces. A difference-in-differences design using the uplift timeline could identify the effect of adding police, complementing the austerity literature on removing them.

7.4 Implications for the Police-Crime Literature

My null result does not mean police have no effect on crime. The literature provides compelling evidence of police deterrence from settings with more credible identification: Draca et al. (2011) on London bus-stop redeployments, Mello (2019) on COPS grants, and Chalfin and McCrary (2018) across multiple natural experiments. My finding is narrower: at the scale of PFA boundaries in England and Wales, the spatial variation in police cuts during austerity does not provide clean identification of the police-crime effect.

This null result is policy-relevant in its own right. The widely-cited correlation between austerity cuts and rising crime—where forces that lost the most officers also saw the largest crime increases—is almost certainly confounded by the underlying geography that determined both the cuts and the crime levels. The BDD reveals that, at shared borders, the "high-cut" forces actually have *lower* crime, underscoring how misleading aggregate cross-force comparisons can be. Cross-sectional comparisons of police forces before and after austerity cannot identify the causal effect of the cuts.

7.5 Limitations

Several limitations warrant discussion. First, the crime data are based on police-recorded incidents, which may be affected by recording practices that vary across forces. Recording cultures can differ persistently between forces—for example, in how anti-social behavior is classified or how proactively incidents are logged. Such persistent differences would generate a stable boundary discontinuity in recorded crime that mirrors the pattern I observe. While this alternative does not change the paper's core conclusion (the BDD does not identify a causal policing effect), it complicates the interpretation of the discontinuity as reflecting "true" crime differences versus measurement differences. Outcomes less sensitive to recording discretion (e.g., hospital admissions, victimization surveys) would provide a useful robustness check.

Second, the signed distance variable is constructed from LSOA centroids derived from crime report coordinates, not official Census centroids. While this introduces some measurement error, the centroid estimates are based on large numbers of geocoded crime reports and closely approximate official centroids.

Third, the analysis is limited to England and Wales and to the specific institutional context of PFA-level policing. The results may not generalize to settings with different police organizational structures or funding mechanisms.

Fourth, the null result applies to the BDD identification strategy, not to the question of whether police affect crime. A difference-in-differences design exploiting within-force temporal variation, or an instrumental variables approach using the grant dependency formula, might yield different conclusions.

Fifth, the workforce data used to define the treatment—percentage change in officer numbers between 2010 and 2018—captures only one dimension of the austerity shock. Forces also lost civilian staff (analysts, call handlers, forensic specialists) and experienced qualitative changes in policing strategy, such as the shift from neighborhood policing to response-only models. If these non-officer reductions affected crime through channels other than deterrence, the officer-based treatment variable may understate the total impact of austerity. Conversely, forces that lost the most officers may have compensated through efficiency gains, technology adoption (CCTV, automated number plate recognition), or partnerships with local authorities and private security. These compensating mechanisms would attenuate any detectable crime effect at boundaries, biasing the BDD estimate toward zero even if police reductions did affect crime. Disentangling the officer headcount channel from these complementary changes requires data on force-level operational changes that are not currently available in standardized form.

Sixth, the LSOA-level analysis may mask heterogeneous effects at finer geographic scales. If police cuts primarily affected crime in specific micro-locations—town centers, transport hubs, nighttime economy areas—these effects could be diluted when aggregated to the LSOA level. A point-level analysis using geocoded crime reports and precise boundary distances could capture more localized effects, though at the cost of increased noise from geocoding imprecision. The tension between spatial resolution and measurement precision is a general challenge in geographic RDD designs (Keele and Titiunik, 2015).

7.6 Broader Implications for Austerity Research

The findings contribute to the growing literature on the consequences of fiscal austerity in the UK. Fetzer (2019) demonstrates that austerity-driven welfare cuts increased support for Brexit; Beatty and Fothergill (2013) documents the spatial concentration of benefit cuts in already-deprived areas. This paper adds a parallel finding for policing: the areas that experienced the deepest police cuts were already the most crime-affected, creating a spurious correlation between cuts and crime levels.

The common thread across these studies is that austerity was not randomly assigned. The distribution of cuts across places was determined by pre-existing fiscal structures—grant dependency for police, benefit receipt rates for welfare, council tax base for local government. These same structures correlate with deprivation, creating endogeneity that threatens any cross-sectional evaluation of austerity's effects. The lesson for future austerity research is that credible identification requires either within-unit temporal variation (difference-in-differences with plausible parallel trends) or exogenous instruments for the size of cuts (formula-based variation that is orthogonal to outcome trends).

The UK provides unusually rich data for studying these questions. The combination of fine-grained geographic crime data (LSOA-level, monthly), detailed police workforce statistics, well-defined administrative boundaries, and extended time series (2011–2024) creates opportunities for research designs that are infeasible in many other contexts. The challenge is not data availability but identification: finding variation in policing that is genuinely exogenous to the conditions that generate crime.

8. Conclusion

I exploit 99 Police Force Area boundaries in England and Wales to test whether differential police cuts during the 2010–2018 austerity period caused crime to diverge at force borders. A boundary discontinuity design reveals a large and significant crime gap at boundaries—but with the "wrong" sign: forces that lost more officers have lower crime at their borders. An event study demonstrates this gap is constant from 2011 to 2024—it neither widens during the cuts nor narrows during the subsequent police uplift. The pre-existing crime discontinuity reflects geographic sorting along administrative boundaries, not a causal effect of reduced policing.

This finding carries two lessons. For policy, it cautions against interpreting the cross-sectional correlation between police cuts and crime as evidence that austerity caused crime to rise. At shared borders, the high-cut forces actually have *lower* crime, underscoring how misleading aggregate comparisons can be. The crime gap reflects historical patterns of urbanization, deprivation, and administrative geography that predate the cuts by decades.

For methodology, this paper demonstrates the value of event study diagnostics within boundary discontinuity designs. Without the temporal dimension, the BDD would have produced a compelling but spurious result—a cautionary tale for applied researchers. The general lesson is that spatial proximity alone is insufficient for identification when boundaries are endogenous. Panel data, where available, provide a crucial diagnostic: if the spatial "effect" is visible before the treatment, the design is compromised.

The question of whether police affect crime remains open in the UK context and is unlikely to be answered by cross-force comparisons alone. Future work might exploit the formula-driven variation in central grants as an instrument for police staffing, use within-force spatial reallocation during austerity to identify neighborhood-level policing effects, or study the staggered Police Uplift Programme using modern difference-in-differences estimators. The crime gap at force boundaries, while not caused by differential policing, provides a valuable mapping of England's geography of deprivation and public safety—a map that policymakers would do well to study.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

References

Beatty, Christina and Steve Fothergill, "Hitting the Poorest Places Hardest: The Local and Regional Impact of Welfare Reform," *Sheffield Hallam University Centre for Regional Economic and Social Research*, 2013.

Becker, Gary S, "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217.

Bell, Brian, Laura Jaitman, and Stephen Machin, "Crime Deterrence: Evidence from the London 2011 Riots," *Economic Journal*, 2014, *124* (576), 480–506.

Bradford, Ben, "Convergence, Not Divergence? Trends and Trajectories in Police-Contact and Confidence in Police," *British Journal of Criminology*, 2011, *51* (1), 179–200.

Calonico, Sebastian, Matias D Cattaneo, and Rocío Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Chalfin, Aaron and Justin McCrary, "Productivity and Selection of Human Capital with Machine Learning," *American Economic Review*, 2018, *108* (5), 1141–1170. See also: Chalfin, Aaron, and Justin McCrary. "Are US cities underpoliced?" *Review of Economics and Statistics*, 100(1), 2018, 167–186.

_ , Benjamin Hansen, Emily K Lerner, and Lucie Parker, "Police Force Size and Civilian Race," *American Economic Review: Insights*, 2022, *4* (2), 139–158.

Dell, Melissa, "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 2010, *78* (6), 1863–1903.

Disney, Richard, Laia Maynou, and Abigail Sheridan, "What Can Neighbourhood Data Tell Us About the Effects of Crime on Local Economic Activity?," Technical Report, University of Nottingham, School of Economics 2019.

Draca, Mirko, Stephen Machin, and Robert Witt, "Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks," *American Economic Review*, 2011, *101* (5), 2157–2181.

Dube, Arindrajit, "Minimum Wages and the Distribution of Family Incomes," *American Economic Journal: Applied Economics*, 2019, *11* (4), 268–304.

Ehrlich, Isaac, "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," *Journal of Political Economy*, 1973, *81* (3), 521–565.

Fetzer, Thiemo, "Did Austerity Cause Brexit?," *American Economic Review*, 2019, *109* (11), 3849–3886.

Haining, Robert P and Guangquan Li, "Ecological Analysis of Crime and Disorder in England and Wales," *Geographical Analysis*, 2014, *46* (1), 53–77.

HM Inspectorate of Constabulary and Fire & Rescue Services, "State of Policing: The Annual Assessment of Policing in England and Wales 2017," Technical Report, HMICFRS 2017.

Keele, Luke J and Rocío Titiunik, "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 2015, *23* (1), 127–155.

Levitt, Steven D, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, 1997, *87* (3), 270–290.

Linden, Leigh and Jonah E Rockoff, "Measuring the Effect of Registries of Sex Offenders on Crime: Evidence from a Natural Experiment," *American Economic Journal: Applied Economics*, 2008, *65* (1), 260–293.

Machin, Stephen, Olivier Marie, and Sunčica Vujić, "Crime and Economic Incentives," *Journal of Human Resources*, 2011, *46* (4), 958–989.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Mello, Steven, "More COPS, Less Crime," *Journal of Public Economics*, 2019, *172*, 174–200.

Nagin, Daniel S, "Deterrence: A Review of the Evidence by a Criminologist for Economists," *Annual Review of Economics*, 2013, *5*, 83–105.

A. Data Appendix

A.1 Crime Data Construction

Crime data are drawn from the data.police.uk bulk archives (`https://data.police.uk/data/archive/`). Each monthly archive contains per-force CSV files with individual crime reports geocoded to LSOA level. I extract data from three cumulative archives:

- 2017-03 archive: contains months from 2011-01 to 2017-03
- 2020-12 archive: contains months from 2018-01 to 2020-12
- 2024-06 archive: contains months from 2021-07 to 2024-06

From these archives, I extract 25 target months. For most years, I sample June and December (two months per year). Three years deviate due to archive boundaries: 2017 uses March only (the last month in the 2017-03 archive); 2021 uses December only (the 2024-06 archive begins in July 2021); and 2024 uses June only (the latest available). Each extraction parses per-force street crime CSVs, aggregates to LSOA $\times$ month $\times$ crime type, and computes LSOA centroids from the median latitude and longitude of all geocoded reports.

The raw data yield 4,084,647 LSOA-month-type observations across 36,705 unique LSOAs and 25 months. After aggregation to the LSOA-year level: 475,105 LSOA-year observations across 14 years (2011–2024).

A.2 Sample Restrictions

The analysis sample includes up to 36,705 LSOAs per year with non-missing PFA assignment and boundary distance. The panel is unbalanced: LSOA-years with zero crime reports are excluded (475,105 of a possible 513,870 LSOA-year cells). For the BDD, I use a fixed bandwidth of 2 km from the boundary for the main specification, yielding an effective sample of approximately 75,000 LSOA-year observations. The event study uses MSE-optimal bandwidths within each year (typically 1–2 km), yielding per-year effective samples of 6,000–8,000.

A.3 Police Workforce Data

Officer headcounts are derived from the Home Office Police Workforce Statistics bulletins. The key variable is the percentage change in officer numbers between 2010 (pre-austerity peak) and 2018 (trough). For the 43 forces, this ranges from -31% (Cleveland) to -0.4% (Surrey), with a median of -12%.

A.4 Geographic Processing

PFA boundary polygons (December 2023 edition) are obtained from the ONS ArcGIS service. Distance computation uses the British National Grid (EPSG:27700) projection to ensure distances are measured in meters. For each of the 36,705 LSOAs, I compute: (1) minimum distance to any PFA boundary, (2) PFA assignment via spatial join, and (3) nearest neighboring PFA via nearest-feature query conditional on excluding the LSOA's own PFA.

Boundary pairs are identified using `st_touches()` on the PFA polygons, yielding 99 pairs. Each near-boundary LSOA is assigned to the pair formed by its own PFA and its nearest neighboring PFA.

B. Identification Appendix

B.1 McCrary Density Test

The McCrary density test (McCrary, 2008) examines whether the density of the running variable is continuous at the cutoff. In this context, it tests whether LSOAs are differentially distributed on either side of PFA boundaries. Since LSOA geographies are fixed, the test is run on the cross-section of unique LSOA centroids in the latest year ($N = 31{,}748$, reflecting only LSOAs with non-zero crime reports in 2024; the full universe of 36,705 LSOAs includes some that report zero crimes in certain years). The test statistic is 0.30 with a p-value of 0.76, using 21,888 observations on the left and 9,860 on the right. There is no evidence of manipulation or sorting.

B.2 Covariate Balance

Using early-period (2011–2012) crime as a balance variable, the RDD yields a coefficient of -0.220 (SE $= 0.037$, $p < 0.001$). This establishes that the crime discontinuity predates the austerity cuts, violating the identifying assumption of the BDD for estimating a causal austerity effect. The IMD (Index of Multiple Deprivation) was unavailable for download during this analysis; future work should incorporate it as an additional balance check.

B.3 Placebo Cutoffs

I estimate the BDD at artificial cutoffs placed at ± 1, ± 2, and ± 3 km from the true PFA boundary:

- -3 km: coefficient $= 0.003$, $p = 0.84$

- -2 km: coefficient $= 0.015$, $p = 0.31$
- -1 km: coefficient $= -0.037$, $p = 0.02$
- $+1$ km: coefficient $= 0.078$, $p < 0.001$
- $+2$ km: coefficient $= 0.079$, $p < 0.001$
- $+3$ km: coefficient $= 0.073$, $p < 0.001$

The asymmetry—significant effects on the positive side but not the negative—is consistent with complex within-force crime gradients. The positive-side placebo coefficients have the opposite sign from the main boundary estimate (-0.20), indicating that the crime landscape within the high-cut forces reverses direction away from the boundary.

C. Robustness Appendix

C.1 Bandwidth Sensitivity

The full bandwidth sensitivity results, using the main specification bandwidth of 2 km as the reference:

Bandwidth multiplier	BW (km)	Coefficient	SE	N_{eff}
0.5× baseline	1.0	−0.167	0.005	37,689
0.75× baseline	1.5	−0.206	0.004	58,318
1.0× (main)	2.0	−0.199	0.003	75,748
1.25× baseline	2.5	−0.182	0.003	90,069
1.5× baseline	3.0	−0.168	0.002	103,267
2.0× baseline	4.0	−0.157	0.002	124,980

C.2 Donut RDD

Because the main specification uses a fixed bandwidth of 2 km, the donut RDD necessarily uses MSE-optimal bandwidth selection (which yields a wider window of approximately 4.5 km) to retain sufficient observations after excluding LSOAs near the boundary. Donuts of 0.5 and 1.0 km failed to converge even with optimal bandwidth. The 2.0 km donut yields a coefficient of 0.071 ($p < 0.05$), reversing sign from the main estimate (-0.199). This sign flip when removing boundary-proximate observations confirms that the discontinuity is a highly local boundary phenomenon driven by the geographic composition of areas immediately adjacent to force borders. The donut result is not directly comparable to the main specification because of the different bandwidth selection method.

C.3 COVID Exclusion

Excluding 2020–2021 yields a coefficient of -0.208 (SE $= 0.003$, $p < 0.001$), nearly identical to the full-sample estimate of -0.199. The COVID pandemic does not drive the results.

C.4 Heterogeneity by Cut Differential

Splitting boundary pairs at the median cut differential shows consistent negative coefficients: pairs with above-median differentials yield -0.21 and below-median pairs yield -0.17. The slightly larger magnitude for high-differential boundaries is in the wrong direction for a deterrence story (the high-cut side has even lower crime where the policing differential is largest), reinforcing the geographic sorting interpretation.

D. Event Study Results

The full event study coefficients:

Year	Coefficient	SE	95% CI lower	95% CI upper
2011	−0.205	0.102	−0.405	−0.006
2012	−0.231	0.102	−0.431	−0.031
2013	−0.210	0.103	−0.413	−0.008
2014	−0.212	0.103	−0.414	−0.010
2015	−0.236	0.099	−0.431	−0.042
2016	−0.200	0.100	−0.396	−0.003
2017	−0.234	0.098	−0.426	−0.042
2018	−0.212	0.097	−0.403	−0.022
2019	−0.209	0.098	−0.401	−0.017
2020	−0.207	0.101	−0.405	−0.008
2021	−0.134	0.075	−0.282	0.013
2022	−0.208	0.086	−0.377	−0.039
2023	−0.229	0.085	−0.395	−0.062
2024	−0.161	0.084	−0.327	0.004

The mean coefficient across all years is -0.207 with a standard deviation of 0.026, confirming temporal stability. A joint test of $\hat{\tau}_t = \bar{\tau}$ for all t fails to reject at the 5% level.