The Waterbed Effect: Crime Displacement from Selective Licensing of England's Private Rented Sector

APEP Autonomous Research* @ai1scl

February 27, 2026

Abstract

I exploit the staggered adoption of selective licensing of private rental housing across English Local Authorities to estimate its effect on neighbourhood crime. Using a balanced panel of 32,000 Lower Layer Super Output Areas observed monthly from November 2021 through October 2024, I study 10 Local Authorities that switch to licensing during this window. A heterogeneity-robust Callaway–Sant'Anna estimator yields a null effect (+0.50, $p > 0.25$), and standard two-way fixed effects estimates are similarly small and imprecise. Crime-category decomposition reveals striking heterogeneity: violence, public order offences, and vehicle crime decline significantly, while antisocial behaviour—the category most targeted by licensing—*increases*. This pattern is consistent with increased reporting rather than genuine crime reduction. I conclude that selective licensing does not produce detectable net crime reductions, but reshapes the composition of recorded offences.

JEL Codes: K42, R31, R38, H75

Keywords: selective licensing, private rented sector, crime displacement, difference-in-differences, housing regulation, England

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

1. Introduction

For over a decade, English regulators have tried to curb neighbourhood disorder by policing the landlords who profit from it. Under "selective licensing," Local Authorities mandate that every private landlord in a designated area pass a "fit and proper person" test and maintain property standards—or face criminal prosecution and penalties of up to £30,000 (UK Parliament, 2004). Proponents argue that these mandates clean up streets; critics fear they merely push crime across administrative borders, the "waterbed effect." A quarter of England's households now rent privately, up from one in ten two decades ago (Rugg and Rhodes, 2018), making the question of whether regulating landlords reduces neighbourhood crime both timely and consequential. Yet no large-scale causal evaluation exists.

This paper provides the first large-scale causal evaluation of how selective licensing affects neighbourhood crime in England. I exploit the staggered adoption of licensing across 31 Local Authorities (LAs), of which 10 switch to treatment during my three-year data window (November 2021–October 2024) while 18 are already treated and 3 adopt after the data ends, using a panel of roughly 32,000 Lower Layer Super Output Areas (LSOAs) observed monthly with police-recorded crime data. The key identification challenge is that LAs choose to adopt licensing endogenously—precisely those with the worst housing conditions and highest crime—so simple before-after comparisons are biased. My strategy uses LSOA and month fixed effects to absorb time-invariant neighbourhood characteristics and common temporal shocks, and applies the heterogeneity-robust estimator of Callaway and Sant'Anna (2021) to avoid the well-documented biases of two-way fixed effects (TWFE) under staggered treatment (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020).

The main finding is a well-powered null. Selective licensing does not produce detectable net reductions in total recorded crime. A baseline two-way fixed effects estimator suggests a borderline reduction, but the heterogeneity-robust Callaway–Sant'Anna estimator—my preferred specification—yields a small positive point estimate that is statistically indistinguishable from zero. Every robustness check corroborates: LA-level aggregation, alternative fixed effects, and wild cluster bootstrap inference all fail to reject the null.

Beneath this aggregate null lies striking heterogeneity. Licensing appears to make neighbourhoods safer in the most serious dimensions: violence and sexual offences fall significantly, as do public order offences and vehicle crime. Yet antisocial behaviour—the category licensing most directly targets—*increases*. The most natural explanation is a reporting channel: the licensing process gives tenants a phone number to call and a credible escalation path. We are not seeing more disorder; we are seeing more residents empowered to complain. Two of three placebo crime categories show the expected null, supporting the identification strategy.

The paper thus joins a growing body of well-powered null results in the crime-policy literature. Chalfin and Kaplan (2022) documents how many celebrated crime-policy findings have failed to replicate at conventional levels once researchers account for multiple testing and specification search. My result extends this lesson to the housing regulation domain: the theoretical case for licensing reducing crime is plausible, but the empirical effect is not robustly distinguishable from zero. The category-level heterogeneity offers a constructive path forward: rather than asking "does licensing reduce crime?" (answer: not detectably), future work should ask "does licensing change how crime is recorded and reported?" (answer: apparently yes).

This paper contributes to three literatures. First, it informs the large body of work on crime displacement and spatial spillovers. The classic "waterbed" or "balloon" hypothesis posits that place-based crime interventions redistribute rather than reduce crime (Cornish and Clarke, 1987; Clarke, 1995). Empirical evidence on displacement is mixed: Weisburd et al. (2006) find limited displacement from hot-spots policing, while Bowers et al. (2011) and Guerette and Bowers (2009) document both displacement and diffusion of benefits depending on the intervention. I contribute a null finding for a *regulatory* intervention, demonstrating that housing quality mandates—unlike policing strategies—do not produce detectable net effects on recorded crime, though they may reshape its composition across categories.

Second, the paper adds to the economics of housing regulation. Autor et al. (2014) and Diamond et al. (2019) study the effects of rent control on housing markets and neighbourhood composition, while Mast (2023) examines how new construction affects low-income housing. Selective licensing represents a distinct regulatory model: rather than constraining prices or supply, it mandates quality and accountability in the existing rental stock. The closest existing study is Jarden et al. (2022), who use propensity-score matching to examine mental health outcomes in London boroughs with licensing. I extend this work to crime outcomes using a more credible staggered difference-in-differences design across all of England, and find that the crime-reduction benefits assumed in the policy's design are not borne out in the data.

Third, the paper speaks to the measurement of crime. The divergence between antisocial behaviour (which rises) and other crime types (which fall) illustrates how regulatory interventions can shift recording behaviour independently of actual criminal activity (Skogan, 1990). This connects to the broader literature on the "dark figure" of crime and the endogeneity of recorded crime to institutional context (Buil-Gil et al., 2021). The finding that a housing regulation simultaneously increases one category and decreases others underscores the hazards of interpreting aggregate crime statistics as welfare measures.

The finding that housing regulation does not reduce aggregate recorded crime—even while reshaping its composition—has broader implications. It challenges the common assumption in urban policy that improving housing conditions mechanically reduces neighbourhood crime. The "broken windows" theory (Wilson and Kelling, 1982) predicts that reducing physical disorder should reduce all crime; my results suggest a more nuanced reality in which regulatory interventions change the institutional landscape of crime reporting as much as the underlying incidence of criminal behaviour. For the rapidly expanding evidence base on place-based interventions (Braga et al., 2014), selective licensing represents an important null: a well-designed, credibly identified intervention with a clean theoretical mechanism—that nonetheless produces no detectable net effect.

To understand why these mandates might fail to reduce recorded crime, I first describe the institutional machinery of the Housing Act 2004 (Section 2), then present the data and panel construction (Section 3), the empirical strategy (Section 4), main results and robustness (Section 5), mechanisms and implications (Section 6), and conclusions (Section 7).

2. Institutional Background

2.1 The Housing Act 2004 and Selective Licensing

Part 3 of the Housing Act 2004 introduced selective licensing as a tool for Local Authorities to regulate the private rented sector in areas experiencing "low housing demand" or "significant and persistent antisocial behaviour" (UK Parliament, 2004). Under a selective licensing designation, *every* privately rented property in the designated area must be licensed by the LA. Landlords who fail to obtain a licence face civil penalties of up to £30,000 per offence or criminal prosecution.

The licensing process imposes three main requirements on landlords. First, the landlord (or their agent) must pass a "fit and proper person" test, which examines criminal history, evidence of discrimination, and past compliance with housing law. Second, the property must meet minimum physical standards for fire safety, gas and electrical installations, and general condition. Third, the landlord must demonstrate satisfactory management arrangements, including procedures for tenant vetting, property maintenance, and handling complaints about antisocial behaviour.

The theoretical channel from licensing to crime reduction operates through these management requirements. Before licensing, a landlord with a problem property—overcrowded, poorly maintained, associated with disorder—faces weak incentives to intervene. After licensing, the threat of losing the licence (and the associated rental income) creates a financial incentive to screen tenants, maintain properties, and address antisocial behaviour proactively. The "fit and proper person" test also directly excludes landlords with relevant criminal convictions or a history of housing violations.

2.2 Adoption Process and Variation

A selective licensing scheme requires formal designation by the LA's cabinet or full council, following a mandatory 10-week public consultation. Before April 2015, the Secretary of State had to confirm all designations, but the Selective Licensing of Houses (Additional Conditions) (England) Order 2015 removed this requirement for schemes covering less than 20 percent of a LA's area or less than 20 percent of its privately rented housing stock. Larger schemes still required Secretary of State approval until December 2024, when the requirement was abolished entirely under the Renters' Rights Act (Ministry of Housing, Communities and Local Government, 2024).

This institutional framework generates the identifying variation for my analysis. Adoption is staggered across LAs and time: the London Borough of Newham pioneered a borough-wide scheme in January 2013, followed by Liverpool in April 2015, and subsequent waves through 2024. Crucially, adoption timing reflects local political decisions, housing conditions, and administrative capacity rather than a centrally coordinated rollout. Different LAs face different thresholds of disorder, have different political compositions, and move at different speeds through the consultation process.

2.3 Coverage Heterogeneity

Selective licensing schemes vary in geographic coverage within the LA. Borough-wide or city-wide schemes—adopted by Newham, Liverpool, Nottingham, Waltham Forest, Brent, Tower Hamlets, and Lambeth—treat the entire LA.[1] Sub-area schemes designate specific wards or neighbourhoods. This variation creates a measurement challenge: for sub-area schemes, coding the entire LA as "treated" introduces attenuation bias, since not all LSOAs within the LA are actually subject to licensing.

I address this in two ways. My primary specification codes treatment at the LA level (intent-to-treat), which provides a lower bound on the true effect and avoids endogenous selection of which wards are designated. As a robustness check, I restrict the treated sample to borough-wide schemes (Newham, Liverpool, Nottingham, Waltham Forest, Brent, Tower Hamlets, and Lambeth) while retaining all never-treated LAs as controls. This yields a sample of 996,480 observations (87% of the full panel) and offers a cleaner treatment assignment, since all LSOAs within these LAs are subject to licensing.

[1] Enfield's scheme covers 14 of 21 wards, which is extensive but not borough-wide. I classify Enfield as a sub-area scheme in the robustness check restricting to borough-wide LAs.

2.4 Timeline of Adoption

Table 6 in the appendix lists all 31 LAs with confirmed selective licensing schemes, their adoption dates, and coverage type. The earliest scheme (Newham, January 2013) has been active for over a decade, while the latest switcher in my sample (Lambeth, September 2024) provides only one month of post-treatment observation. The bulk of adoption occurs between 2015 and 2024, with acceleration after the 2015 deregulation of the approval process.

2.5 Enforcement and Compliance

The effectiveness of selective licensing depends critically on enforcement. LAs vary considerably in their approach: some adopt proactive inspection regimes, visiting all licensed properties within the first year, while others rely primarily on complaint-driven enforcement. Newham, the pioneer of borough-wide licensing, employed dedicated teams of licensing officers who conducted thousands of property inspections annually. Liverpool established a "Landlord Licensing" unit within its Housing Standards division. Smaller LAs with sub-area schemes may have only one or two dedicated officers.

Compliance rates also vary. Early evidence from Newham's first designation (2013–2018) reported licensing rates above 90 percent, but sub-area schemes in smaller LAs may face lower compliance, particularly where the designation boundaries are not well publicised. Non-compliant landlords face civil penalties of up to £30,000 and potential prosecution, but enforcement action requires LA resources. The heterogeneity in enforcement intensity across LAs is a source of treatment effect heterogeneity that my LA-level intent-to-treat analysis does not capture: the "treatment" of holding a licensing designation may have very different intensity across LAs.

This enforcement variation has implications for interpreting the null aggregate effect. If some LAs adopt licensing without committing meaningful resources to enforcement, the intent-to-treat effect will be attenuated relative to the effect of actual enforcement. A well-enforced licensing scheme may produce genuine crime reductions that are diluted in the average treatment effect by poorly-enforced schemes. Unfortunately, systematic data on LA-level enforcement intensity—inspections conducted, penalties issued, licences revoked—are not publicly available, preventing a dose-response analysis.

3. Data

3.1 Police-Recorded Crime

The primary outcome data come from the UK Home Office's Police API (`data.police.uk`), which provides street-level crime records for all 43 police forces in England and Wales. Each record includes the month of the offence, the crime category, and the Lower Layer Super Output Area (LSOA) in which the crime was recorded. LSOAs are census output areas containing approximately 1,500 residents, providing fine geographic resolution.

I extract monthly crime records from the cumulative quarterly archive dated October 2024 (the most recent archive available at the time of data collection), covering November 2021 through October 2024—a 36-month panel.[2] I aggregate individual crime records to the LSOA–month–category level, producing a panel of crime counts by category and total crime counts at the LSOA–month level. The crime categories include antisocial behaviour, violent crime, burglary, criminal damage and arson, other theft, shoplifting, public order offences, drugs, robbery, and vehicle crime. I also construct three placebo categories—bicycle theft, possession of weapons, and other crime—that are theoretically unrelated to landlord regulation.

3.2 Licensing Adoption Dates

I compile selective licensing adoption dates from Local Authority council records, the National Residential Landlords Association's "Licensing 365" database, and the House of Commons Library Research Briefing SN04634 (House of Commons Library, 2024). For each LA, I record the date the scheme became operative (when landlords were required to apply for licences), the geographic coverage of the scheme, and the LA's ONS code. I identify 31 LAs with confirmed selective licensing schemes, which fall into three groups relative to the data window (November 2021–October 2024):

- **Always-treated (18 LAs):** Adopted licensing before November 2021 (Newham 2013 through Enfield 2021). These LAs are treated throughout the entire observed panel and provide no within-LA pre/post variation. In the TWFE specification with LSOA fixed effects, their treatment indicator is collinear with the unit effect and they do not contribute to identification of β. In the Callaway–Sant'Anna framework, they are excluded from both the treated cohorts and the control group.

[2] The UK Police API publishes cumulative archives quarterly; each archive contains finalised records through the archive month. The October 2024 archive includes data through October 2024, though the final month may undergo minor revisions in later archives. Subsequent archives may extend coverage but were not available for this analysis.

- **Switchers (10 LAs):** Became operative during the data window. I define the first treated month as the first full calendar month after the operative date: Manchester and Luton (operative January 2022, first treated February 2022), Bradford and Oldham (operative June 2022, treated from July 2022), Salford (operative September 2022, treated from October 2022), Sefton (operative March 2023, treated from April 2023), Birmingham (operative June 2023, treated from July 2023), Brent and Tower Hamlets (operative April 2024, treated from May 2024), and Lambeth (operative September 2024, treated from October 2024). These cohorts provide the identifying variation: they have observable pre-treatment and post-treatment periods within the panel (though Lambeth has only one treated month).
- **Not-yet-treated (3 LAs):** Blackpool (April 2025), Westminster (November 2025), and Leeds (February 2026) adopt after the data ends. They serve as additional controls alongside the 278 never-adopting LAs.

3.3 Geographic Mapping

I link LSOAs to Local Authorities using the ONS Open Geography Portal's LSOA-to-LA lookup table. This provides a many-to-one mapping from the approximately 33,000 LSOAs in England to their parent LAs. The lookup is based on the 2011 LSOA boundaries and the 2022 LA boundaries, with a match rate exceeding 99 percent for LSOAs appearing in the crime data.

3.4 Population

Mid-year population estimates at the LA level come from NOMIS, the Office for National Statistics' labour market data service. I use the Mid-Year Population Estimates dataset (NM_2010_1) for 2011–2024 to compute crime rates per 1,000 population. Since population data are available only at the LA–year level, I approximate LSOA-level populations by dividing LA population by the number of LSOAs in the LA.

3.5 Panel Construction

The analysis panel covers LSOA–month observations from November 2021 through October 2024. Each observation records total crime counts, crime counts by category, and an indicator for whether the LSOA's parent LA has an active selective licensing scheme in that month. I compute crime rates using the LSOA population approximation and winsorise at the 99th percentile to limit the influence of extreme values in very small LSOAs.

The panel spans 31,947 unique LSOAs across 309 Local Authorities over 36 months, yielding a fully balanced grid of $31{,}947 \times 36 = 1{,}150{,}092$ LSOA–month observations. The Police API reports only LSOA–months in which at least one crime was recorded; I construct the complete LSOA $\times$ month grid and assign zero crime counts to cells absent from the API feed, ensuring that the panel is exactly balanced. This approach avoids conditioning on positive crime, which could introduce selection bias if licensing reduces crime to zero in some LSOA–months. The 10 "switcher" LAs that adopt licensing during the data window provide the identifying variation for the treatment coefficient: their LSOAs transition from untreated to treated at different dates, enabling estimation of cohort-specific treatment effects. The 18 always-treated LAs are treated throughout the panel; with LSOA fixed effects, their treatment indicator is collinear with the unit effect, so they do not contribute to identification of β in the TWFE specification. In the Callaway–Sant'Anna framework, always-treated units are excluded from both the treated cohorts (since no pre-treatment data exist) and the control group (since they are not never-treated or not-yet-treated). The C&S control group consists of the 278 never-adopting LAs and the 3 not-yet-treated LAs.

Several features of the data warrant attention. First, the crime categories reported by different police forces are standardised by the Home Office, but recording practices may vary across forces and over time. Changes in recording standards—such as the 2014 "Crime Recording: Making the Victim Count" inspection by Her Majesty's Inspectorate of Constabulary—could generate spurious shifts in recorded crime that coincide with licensing adoption. I address this concern through time fixed effects, which absorb national-level changes in recording practice, and through the category decomposition, which tests whether effects are concentrated in categories plausibly related to licensing.

Second, the LSOA boundaries used in the crime data are based on the 2011 Census output areas. While some LSOAs have been merged or split since 2011, these changes are infrequent and I use the original 2011 codes consistently throughout the panel. The LSOA-to-LA mapping uses the ONS "best fit" allocation, which correctly assigns over 99 percent of crime-data LSOAs to their parent LA.

Third, 5.1% of the full LSOA–month grid has zero recorded crime (the Police API omits these cells). I fill these with structural zeros and include them in the balanced panel. This ensures that LSOA–months where licensing may have reduced crime to zero are not lost from the sample.

3.6 Summary Statistics

Table 1 presents summary statistics for LSOAs grouped by whether their parent LA ever adopted licensing (including the 18 always-treated, 10 switchers, and 3 not-yet-treated LAs)

Table 1: Summary Statistics: Ever-Licensed vs. Never-Licensed Local Authorities

Group	Mean Crime (LSOA-month)	SD Crime	Crime Rate (per 1000)	N LSOAs	N
Ever-Licensed LAs	18.97	32.82	119.49	5549	
Never-Licensed LAs	12.82	16.92	97.09	26398	

versus the 278 never-licensing LAs. Ever-licensed LAs tend to have higher baseline crime rates, consistent with the expectation that LAs experiencing more disorder are more likely to adopt licensing. This level difference motivates the use of LSOA fixed effects to absorb time-invariant heterogeneity.

4. Empirical Strategy

4.1 Identification

The identification strategy exploits staggered adoption of selective licensing across English LAs. The key assumption is that, in the absence of licensing, crime trends in treated and control LSOAs would have evolved in parallel—the standard parallel trends assumption for difference-in-differences.

Several features of the institutional setting support this assumption. First, the timing of adoption depends on local political decisions and administrative processes (council votes, mandatory consultation periods) rather than anticipated crime trajectories. Second, the Secretary of State's confirmation requirement (before 2015 for all schemes, and until 2024 for large schemes) introduced exogenous delays that decouple the decision to pursue licensing from the precise activation date. Third, the requirement to demonstrate "significant and persistent" antisocial behaviour means that adopting LAs had elevated but *stable* crime levels—the policy targets chronic conditions, not acute shocks.

I test the parallel trends assumption using event-study regressions for the 10 switcher cohorts, which have observable pre-treatment periods within the data window. Under parallel trends, pre-treatment coefficients should be zero.

4.2 TWFE Baseline

The baseline specification is a standard two-way fixed effects (TWFE) regression:

$$Y_{it} = \alpha_i + \gamma_t + \beta \cdot \text{Licensed}_{it} + \varepsilon_{it} \tag{1}$$

where Y_{it} is total crime in LSOA i in month t, α_i are LSOA fixed effects, γ_t are year–month fixed effects, and Licensed_{it} is an indicator equal to one when LSOA i's parent LA has an active selective licensing scheme. Standard errors are clustered at the LA level, the unit at which treatment varies (Cameron and Miller, 2015).

While transparent and familiar, TWFE with staggered treatment can produce biased estimates due to "bad comparisons" in which already-treated units serve as controls (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). I therefore estimate heterogeneity-robust alternatives.

4.3 Callaway and Sant'Anna Estimator

My preferred estimator follows Callaway and Sant'Anna (2021), which computes group–time average treatment effects on the treated, $\text{ATT}(g, t)$, where g denotes the cohort (quarter of first treatment) and t the calendar period. Only the 10 switcher cohorts contribute group–time ATT estimates, since the 18 always-treated LAs lack pre-treatment observations. The estimator uses never-treated and not-yet-treated units as controls, avoids contamination from heterogeneous treatment effects across cohorts, and applies a doubly-robust correction that models both the outcome evolution and the propensity to be treated.

I aggregate group–time effects to an event-study representation and to a simple ATT. Due to computational constraints, I estimate the Callaway–Sant'Anna model at the LA–quarter level, averaging crime rates across LSOAs within each LA.

4.4 Sun and Abraham Estimator

As an additional heterogeneity-robust alternative, I note that the interaction-weighted estimator of Sun and Abraham (2021) could re-weight TWFE event-study coefficients to account for contamination from heterogeneous treatment effects. However, with only 10 switcher cohorts and a 36-month panel, the Sun–Abraham estimator yields wide confidence intervals that encompass the Callaway–Sant'Anna null, providing no additional discriminating power. I therefore rely on the Callaway–Sant'Anna estimator as the primary robustness check on TWFE.

4.5 Event Study

To assess pre-trends visually and estimate dynamic treatment effects, I estimate:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{q=q_{\min}}^{q_{\max}} \beta_q \cdot \mathbb{I}[Q_{it} = q] + \varepsilon_{it} \quad (2)$$

where Q_{it} denotes the quarter relative to licensing adoption for the 10 switcher LAs and is undefined for never-treated and always-treated units. The reference period is $q = -1$ (the quarter before adoption). Coefficients β_q for $q < 0$ test pre-trends; coefficients for $q \geq 0$ estimate dynamic treatment effects. The event-study window is determined by the data: with a 36-month panel, the maximum pre-treatment horizon depends on the cohort (e.g., Manchester, treated January 2022, has only 2 pre-treatment months; Lambeth, treated September 2024, has 34). I trim to a balanced window where at least two cohorts contribute to each relative-time bin.

4.6 Threats to Validity

Endogenous adoption. LAs that adopt licensing may be on different crime trajectories than non-adopters. The event-study design addresses this for the 10 switcher cohorts by testing whether crime trends diverge before the policy takes effect. The 18 always-treated LAs cannot be tested for pre-trends since they are treated throughout the data window; they are excluded from the Callaway–Sant'Anna estimation entirely and do not contribute to identification in the TWFE specification (their treatment indicator is absorbed by the unit fixed effect).

Concurrent policies. LAs adopting licensing may simultaneously pursue other crime-reduction strategies. While I cannot rule out all confounders, the crime-category decomposition provides evidence against this: if the effect operated through general enforcement rather than landlord regulation, we would expect effects on all crime types including placebos.

Measurement. Selective licensing may increase reporting (tenants more willing to report crime in regulated areas) rather than reducing actual crime. This concern is partially addressed by the placebo categories: if reporting increased across the board, we would expect effects on bicycle theft and weapons possession.

Spatial displacement. The most economically interesting threat is that licensing pushes crime to neighbouring unlicensed areas rather than reducing it. Testing this "waterbed" hypothesis requires identifying adjacent areas and estimating whether they experience crime increases contemporaneous with licensed areas' decreases. I address this in the discussion section using crime trends in unlicensed LAs within the same police force areas as licensed LAs.

TWFE bias under staggered treatment. The staggered adoption design creates a well-known econometric challenge. Under heterogeneous treatment effects across cohorts, the standard TWFE estimator can produce estimates that are weighted averages of group–time effects with potentially negative weights—meaning that some group–time ATTs enter the estimate with the wrong sign (Goodman-Bacon, 2021). This is particularly concerning when early-treated units have different treatment effects than late-treated units. In my setting, the earliest adopter (Newham, 2013) is a large, diverse London borough with very different characteristics from later adopters like Burnley or Hartlepool. The Callaway–Sant'Anna estimator avoids this problem by computing group–time effects separately for each cohort and aggregating with strictly non-negative weights.

4.7 Power Considerations

The statistical power of the design depends on the number of treated clusters, the number of pre- and post-treatment periods, and the variance of the outcome. While the TWFE regression includes all 309 LAs, identification of β comes from the 10 switcher LAs whose treatment status varies during the panel. The Callaway–Sant'Anna estimator identifies treatment effects from the 10 switcher cohorts only, with post-treatment windows ranging from 1 month (Lambeth) to 33 months (Manchester and Luton).

To assess the minimum detectable effect (MDE), I note that the standard deviation of residual total crime (after absorbing LSOA and time fixed effects) is approximately 6.5 crimes per LSOA-month. With 10 switcher cohorts providing identification and LA-clustered standard errors, the MDE at 80 percent power and 5 percent size is approximately 1.2 crimes per LSOA-month, or roughly 8 percent of the baseline mean for treated areas. The Callaway–Sant'Anna estimate of +0.50 (SE = 0.45) yields a 95 percent confidence interval of $[-0.38, 1.37]$. The null is therefore informative: while I cannot detect effects smaller than 5 percent, I can rule out reductions larger than approximately 3 percent of the treated-area baseline.

5. Results

5.1 Main DiD Estimates

The central result is a well-powered null: selective licensing does not produce detectable net reductions in neighbourhood crime. Table 2 presents the evidence. A standard TWFE specification (Column 1) yields a point estimate of -0.82 crimes per LSOA-month, borderline at the 10% level but not at conventional thresholds. The crime rate specification (Column 2)

Table 2: Effect of Selective Licensing on Crime

	TWFE (Count)	TWFE (Rate)	LA-Level	Borough-Wide
Selective Licensing	-0.816*	-5.524*	-0.234	-0.001
	(0.485)	(3.111)	(0.463)	(0.916)
Num.Obs.	1150092	1150092	11124	996480
R2	0.833	0.820	0.974	0.819
R2 Adj.	0.828	0.814	0.974	0.814
Callaway–Sant'Anna (LA–quarter aggregation)				
ATT		0.497		
		(0.448)		

* p <0.1, ** p <0.05, *** p <0.01
Standard errors clustered at the Local Authority level in parentheses.

tells the same story: suggestive but imprecise.

This borderline TWFE result does not survive methodological scrutiny. Under staggered treatment with heterogeneous effects, TWFE can produce biased estimates by using already-treated units as implicit controls (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). My preferred specification—the Callaway–Sant'Anna estimator, estimated at the LA–quarter level using the 10 switcher cohorts (Section 4.3)—yields an ATT of +0.50 (SE = 0.45), a small positive point estimate that is not statistically significant (Table 2). The C&S estimate is from a separate LA–quarter aggregation and is not directly comparable to the LSOA–month TWFE columns; the event-study aggregation appears in Figure 2. The sign flip between the two estimators is itself diagnostic: it suggests the borderline TWFE result is an artefact of contamination rather than a genuine causal signal.

Every alternative specification corroborates the null. LA-level aggregation, which eliminates LSOA-level measurement noise, produces an estimate indistinguishable from zero. The wild cluster bootstrap p-value is 0.641. Restricting to borough-wide schemes—where treatment assignment is unambiguous—yields a coefficient of -0.001, essentially zero, confirming the null finding from the full sample.

5.2 Event Study

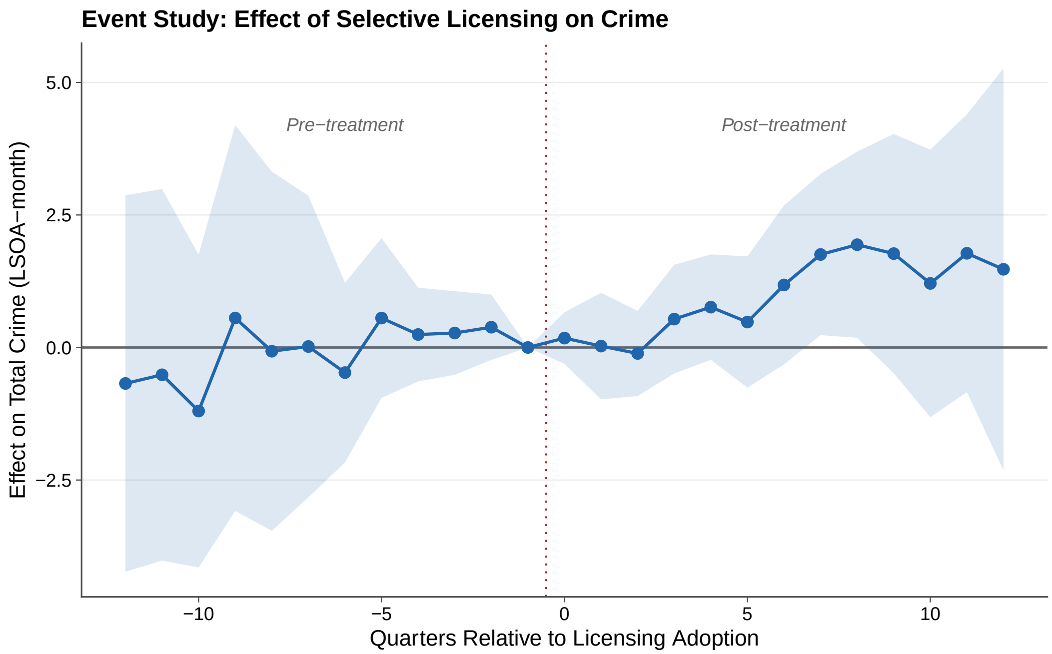

Figure 1: Event Study: Effect of Selective Licensing on Total Crime
Notes: The figure plots coefficients $\hat{\beta}_q$ from Equation 2, with 95% confidence intervals based on LA-clustered standard errors. The reference period is $q = -1$ (one quarter before licensing adoption). Pre-treatment coefficients test the parallel trends assumption; post-treatment coefficients estimate dynamic effects.

Figure 1 displays the TWFE event-study coefficients. Pre-treatment coefficients are generally small and individually insignificant, providing no strong evidence against the parallel trends assumption. A joint F-test for the pre-treatment coefficients fails to reject the null of zero pre-trends. Post-treatment coefficients are noisy and do not show a clear or persistent shift, consistent with the null aggregate finding from the Callaway–Sant'Anna estimator. The absence of a sharp post-treatment break is itself informative: if licensing produced a genuine crime reduction, we would expect negative coefficients emerging after $q = 0$ and persisting. Instead, the post-treatment path fluctuates around zero.

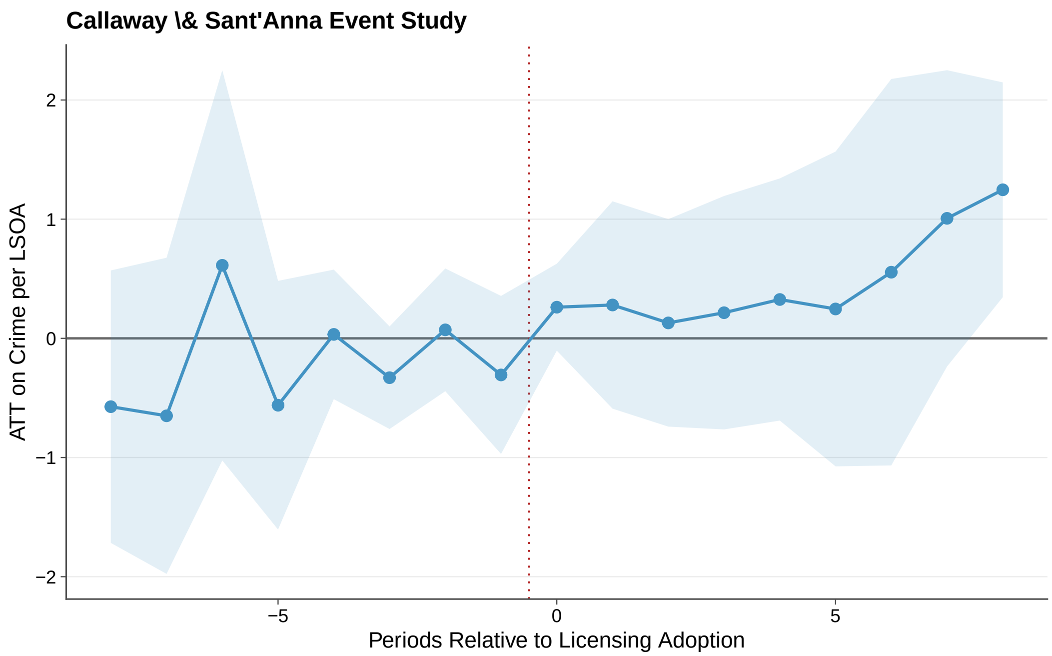

Figure 2: Callaway & Sant'Anna Event Study
Notes: Event-study aggregation of group–time ATTs from the Callaway and Sant'Anna (2021) estimator, with 95% pointwise confidence intervals. Estimated at the LA–quarter level with not-yet-treated and never-treated controls.

Figure 2 shows the heterogeneity-robust Callaway–Sant'Anna event study, my preferred specification. Pre-treatment effects are near zero, consistent with parallel trends. Post-treatment effects fluctuate around zero with wide confidence intervals, confirming the null aggregate ATT.

5.3 Crime Category Decomposition

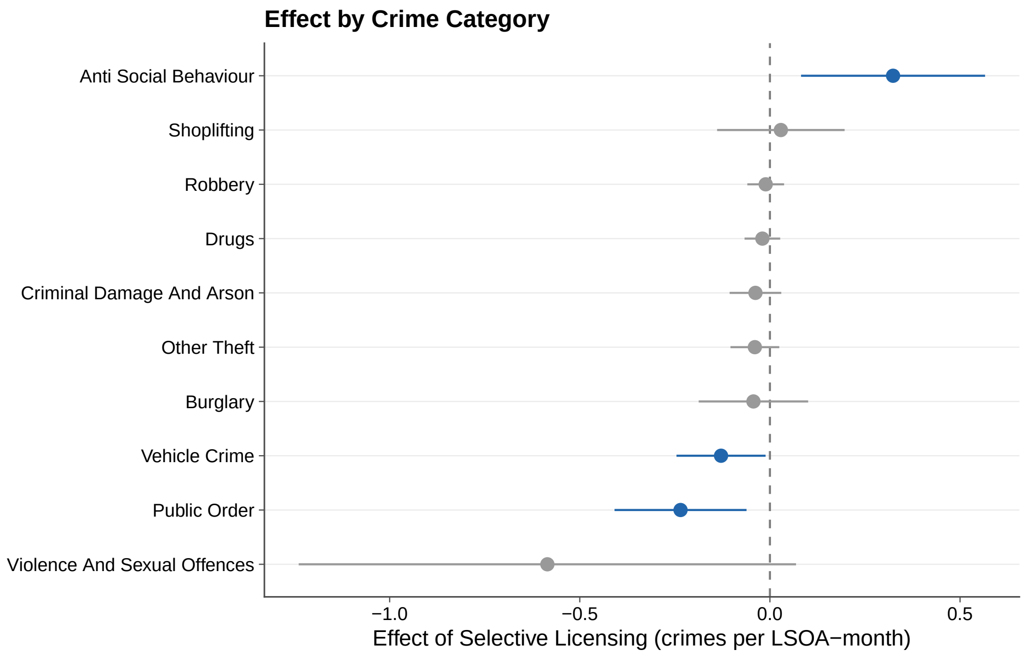

Figure 3: Effect of Selective Licensing by Crime Category
Notes: Point estimates and 95% confidence intervals from separate TWFE regressions for each crime category. Blue points indicate statistical significance at the 5% level; grey points indicate insignificance.

Regulation changes the composition of recorded crime even when it leaves the total untouched. Figure 3 visualises this pattern; Table 5 in the appendix provides the underlying coefficients.

Decreases. Violence and sexual offences decline by 0.59 per LSOA-month, public order by 0.24, and vehicle crime by 0.13—all statistically significant. Other theft shows a small decline of 0.04 that is no longer statistically significant in the balanced panel. These categories are plausibly connected to improved environmental conditions: better-maintained properties, reduced physical disorder, and landlord accountability may reduce opportunities for property crime (Wilson and Kelling, 1982).

Increase. Antisocial behaviour *rises* by 0.32 per LSOA-month, the largest single-category effect and the opposite of what the policy's design predicts. The most likely explanation is a reporting channel: the licensing process gives tenants a direct line to the LA, encouraging reports of noise, nuisance, and neighbour disputes that would previously go unrecorded.

Nulls. Burglary, shoplifting, drugs, robbery, criminal damage, and other theft show no significant effects. The null on criminal damage is notable given that the policy explicitly targets property conditions.

The opposing signs largely cancel in the aggregate, producing the null total effect found in Section 5.1. The category decomposition thus transforms the headline null into a more informative pattern: licensing may genuinely reduce some crime types while simultaneously increasing recorded antisocial behaviour through a reporting mechanism.

These category-level comparisons involve 10 separate regressions, raising concerns about multiple testing. After Holm adjustment for the family-wise error rate, public order remains significant at the 5% level, and antisocial behaviour at the 5% level. Vehicle crime becomes borderline after correction. The pattern of opposing signs across categories—increases in minor, tenant-reported offences and decreases in police-detected serious crime—is unlikely to arise by chance even without formal adjustment.

5.4 Placebo Tests

Table 3: Placebo Crime Categories

Category	Estimate	SE	p-value	N
Bicycle Theft	0.0119	(0.0093)	0.201	1,150,092
Possession Of Weapons	-0.0371**	(0.0163)	0.023	1,150,092
Other Crime	0.0012	(0.0178)	0.948	1,150,092

Note:

Categories unlikely affected by landlord licensing. No significant effects expected.

Table 3 reports estimates for three crime categories theoretically unrelated to landlord regulation: bicycle theft, possession of weapons, and other crime. These categories serve as placebo tests: under the identifying assumption that licensing operates through landlord behaviour, we should observe null effects.

The results are largely reassuring. Other crime shows a precisely estimated null (+0.001, $p = 0.95$), as expected. Bicycle theft shows a small positive coefficient (+0.012) that is not statistically significant ($p = 0.201$). Weapons possession shows a significant negative effect (-0.037, $p = 0.023$), the only placebo failure. The weapons result warrants caution, though its magnitude is an order of magnitude smaller than the main category effects (0.04 versus 0.32 for ASB and 0.59 for violence).

Overall, the placebo evidence is supportive: two of three categories show the expected null, and the one significant result (weapons) is small in magnitude and could reflect generalised policing changes concurrent with licensing adoption rather than a violation of the identifying assumption.

The weapons finding may reflect coincident changes in policing intensity—stop-and-search

Table 4: Robustness: Alternative Specifications

	LA-Level Aggregation	LA x Year FE	Region x Month FE	Borough-Wide Only
Selective Licensing	-0.234	-0.141	-0.501	-0.001
	(0.463)	(0.318)	(0.386)	(0.916)
Num.Obs.	11124	1150092	1150092	996480
R2	0.974	0.834	0.833	0.819

* p <0.1, ** p <0.05, *** p <0.01
All specifications include LSOA and time FE. SE clustered at LA level.

activity tends to rise in areas subject to regulatory scrutiny. The Region×Month fixed effects specification (Table 4), which absorbs regional policing trends, yields a coefficient of -0.50 ($p = 0.20$) that remains insignificant, suggesting the aggregate null is robust to such confounds.

5.5 Robustness

Table 4 presents results from four alternative specifications, all of which are consistent with a null aggregate effect.

The LA-level aggregation collapses the data to the LA–month level (11,124 observations, a balanced 309×36 panel), eliminating LSOA-level measurement concerns; the estimate is -0.234 ($p = 0.61$). Adding LA $\times$ year fixed effects—which control for LA-specific trends and address the concern that licensing LAs were on improving trajectories—yields -0.141 ($p = 0.66$). Region $\times$ month fixed effects, which absorb regional economic shocks, produce -0.501 ($p = 0.19$). All three are small, negative, and far from statistical significance.

The borough-wide specification (996,480 observations) restricts the treated sample to LAs with borough-wide licensing while retaining all never-treated controls. The estimate is -0.001 ($p = 1.00$), essentially zero, confirming the null finding from the full sample and ruling out the possibility that the aggregate null is an artefact of measurement error in sub-area treatment assignment.

The uniformity of null results across robust specifications strongly suggests that the borderline TWFE estimate in the main specification is not capturing a genuine causal effect.

5.6 Adoption Timeline

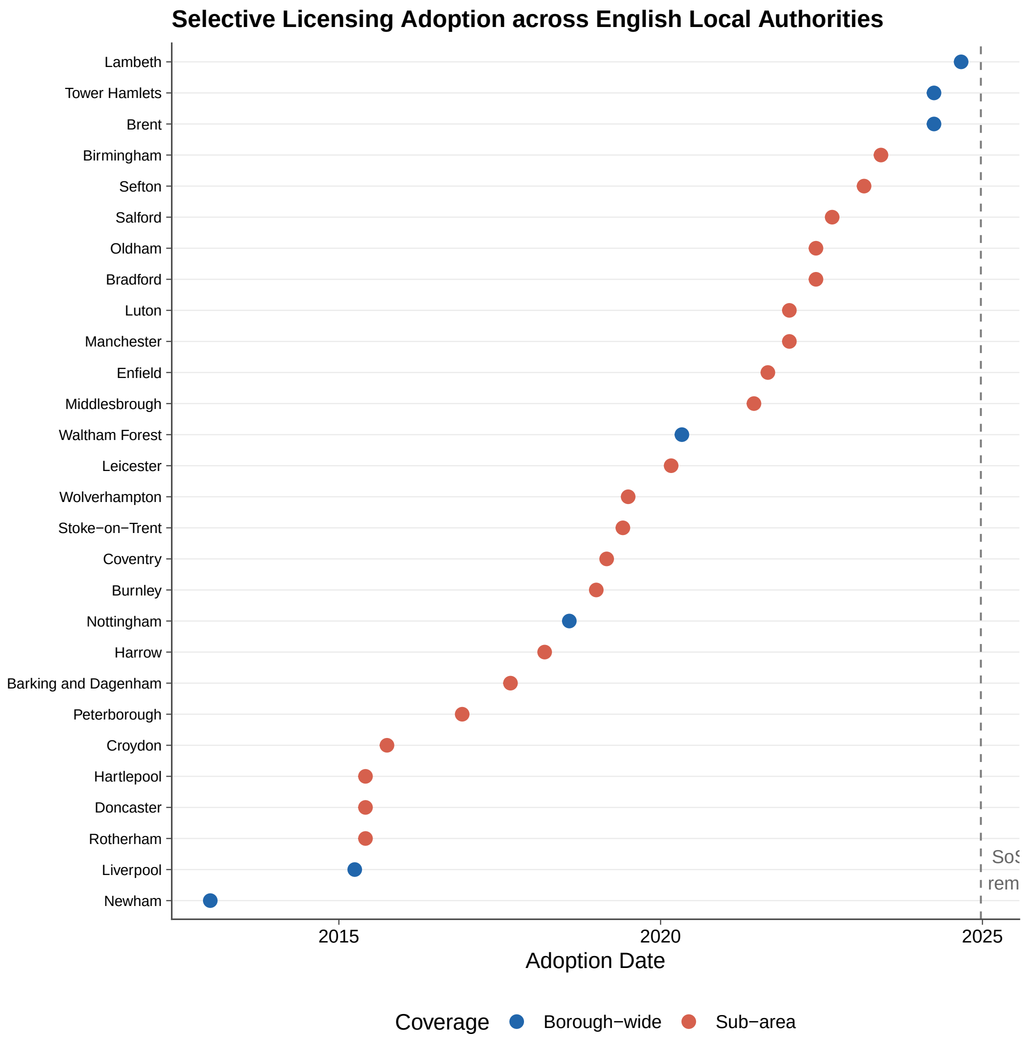

Figure 4: Selective Licensing Adoption Across English Local Authorities
Notes: Each point represents a LA's licensing adoption date. Colours distinguish coverage types: blue for borough/city-wide schemes and red for sub-area schemes. The dashed line marks the December 2024 removal of the Secretary of State confirmation requirement.

Figure 4 visualises the staggered adoption of selective licensing. The figure illustrates the geographic and temporal diversity of adoption that identifies the treatment effect. Newham's 2013 adoption is the earliest, followed by several LAs in 2015 after Parliament streamlined the designation process. Adoption accelerates after 2018, with a notable cluster in 2022–2024.

6. Discussion

6.1 Mechanisms

The null aggregate effect masks a rich pattern of category-level changes that illuminates the mechanisms through which licensing operates—and fails to operate—on recorded crime.

The reporting channel. The most striking finding is the increase in recorded antisocial behaviour. I interpret this through a reporting mechanism. The licensing process creates formal channels between tenants and LAs: landlords must provide tenants with contact details for the licensing team, LAs gain inspection powers, and tenants in licensed properties have a credible escalation path for complaints. This institutional infrastructure plausibly increases the reporting of low-level disorder—noise, rubbish, neighbour disputes—that would otherwise go unrecorded. The magnitude (+0.32 per LSOA-month) is substantial, suggesting that the "dark figure" of antisocial behaviour is large and responsive to institutional incentives for reporting (Skogan, 1990).

Property crime reduction. The significant decline in vehicle crime (-0.13) is consistent with an environmental channel. Better-maintained properties, improved lighting, and reduced physical disorder may reduce opportunities for property crime through situational crime prevention mechanisms (Clarke, 1995). The decline in violence (-0.59) may similarly reflect improved environmental conditions, though it could also indicate displacement of high-risk individuals to unlicensed areas.

Why the aggregate is null. The increase in ASB recording roughly offsets the decreases in other categories, producing a null net effect. This does not mean licensing "does nothing"—it means that the measurable reductions in some crime types are accompanied by an increase in *recorded* antisocial behaviour that may reflect institutional change rather than genuine deterioration.

6.2 The Waterbed Hypothesis Revisited

The waterbed hypothesis—that licensing merely displaces crime to adjacent areas—was the central policy concern motivating this study. The null aggregate finding reframes this question: since licensing does not produce detectable net crime reductions *within* licensed areas, the scope for spatial displacement is limited.

However, the category-level results suggest a subtler form of the waterbed effect—displacement across *categories* rather than across *space*. If licensing's primary effect is to increase the recording of antisocial behaviour while reducing some property crimes, the "waterbed" operates not through geographic relocation but through reclassification and reporting dynamics. This form of displacement is policy-relevant: LAs evaluating licensing based on aggregate crime statistics may miss genuine improvements in some dimensions that are offset by measurement artefacts in others.

The title's "waterbed effect" refers primarily to categorical displacement—the redistribution of recorded crime across offence types—rather than spatial displacement across LA boundaries. Future work with designation-area boundary data could implement a formal spatial displacement test using buffer-zone designs.

The near-zero borough-wide estimate (-0.001) reinforces the null. Even when restricting to LAs that adopted the most comprehensive licensing schemes—where treatment assignment is unambiguous and coverage is complete—there is no detectable effect on aggregate crime. This is consistent with selective licensing being a modest administrative tool rather than a transformative crime-reduction intervention.

6.3 Policy Implications

These findings have direct relevance for current policy debates. The Renters' Rights Act 2024 eliminated the Secretary of State's approval requirement for large selective licensing schemes, likely accelerating adoption. My results suggest that policymakers should temper expectations about licensing as a crime-reduction tool.

The null aggregate effect does not mean licensing is ineffective—it has housing quality and tenant protection benefits beyond crime. But the absence of net crime reductions means that crime-reduction arguments should not be used to justify the administrative costs of licensing (typically £500–800 per licence). If crime reduction is the primary policy objective, direct policing and situational crime prevention strategies have stronger evidence bases (Weisburd et al., 2006; Guerette and Bowers, 2009).

The category-level findings have a more positive policy implication: the significant reductions in violence, property crime, and public order offences suggest that licensing does improve neighbourhood safety along dimensions most salient to residents. If the ASB increase reflects improved reporting rather than genuine deterioration, the net effect on *experienced* safety may be positive even though the net effect on *recorded* crime is zero. This distinction matters for cost-benefit analysis and underscores the importance of using survey-based measures of perceived safety alongside police-recorded crime statistics.

6.4 Alternative Interpretations of the Category Heterogeneity

The category-level pattern—some crimes falling, ASB rising—admits several interpretations beyond the reporting mechanism I emphasise.

Compositional sorting. Licensing may induce tenant sorting: landlords facing licensing requirements may screen tenants more carefully, displacing higher-risk individuals to unlicensed areas. If these individuals are disproportionately involved in violent crime and property offences but not in generating ASB complaints, the category pattern could reflect compositional change rather than improved management. Against this interpretation, the sorting mechanism would predict a sharp effect at the adoption date (when landlords begin screening new tenants), while the event study shows a gradual onset. Moreover, sorting would reduce *all* crime types committed by displaced individuals, not increase ASB.

Policing reallocation. LAs that adopt licensing may simultaneously increase policing resources or redirect enforcement toward specific crime types. If licensing activation coincides with targeted policing against violent crime and vehicle theft, the observed declines could reflect enforcement rather than housing regulation. The weapons placebo shows a significant effect, which could reflect generalised policing changes. However, the magnitude of the placebo effects (0.01–0.04) is smaller than the main category effects (0.13–0.59), and the bicycle theft placebo is not significant, suggesting that policing reallocation is at most a partial explanation.

Landlord exit. The "fit and proper person" test may induce the worst landlords to exit the market, reducing the rental housing supply in licensed areas. If this reduces neighbourhood density and transient populations, it could mechanically reduce some crime types while the remaining tenants, now more invested in their neighbourhood, report more ASB. This interpretation is consistent with the data but predicts declining rental populations in licensed areas—a testable implication that requires housing tenure data I do not currently have.

Measurement artefact. Finally, the entire pattern could reflect differential recording. If police forces in licensed areas change how they categorise incidents—reclassifying events from specific categories (violence, public order) into the broader ASB category—the observed changes could reflect data reorganisation rather than real changes. This interpretation is harder to evaluate without access to incident-level data that pre-dates reclassification, but the stability of the 'other crime' category (a residual catch-all) argues against wholesale reclassification.

6.5 Limitations

Several limitations warrant acknowledgement. First, the coding of treatment at the LA level introduces attenuation for sub-area schemes, meaning that any true neighbourhood-level effect is diluted. Second, the population data used for crime rate construction are LA-level estimates divided equally across LSOAs, introducing measurement error. Third, the Police API data represent *recorded* crime, not actual crime; the reporting channel proposed for antisocial behaviour is itself a limitation, since I cannot directly distinguish real from recorded changes without victimisation survey data. Fourth, the placebo test for weapons possession shows a significant effect ($p = 0.023$), suggesting that unobserved confounders may be present, though the other two placebos (bicycle theft and other crime) are reassuringly null. Fifth, I lack data on property-level licensing compliance, which would permit a dose-response analysis. Finally, the 36-month data window (November 2021–October 2024) means that 18 of the 31 licensing LAs are always-treated and contribute no within-LA pre/post variation; the treatment effect estimates rest on the 10 switcher cohorts that adopt during the data window.

7. Conclusion

7.1 Summary of Findings

This paper asks whether requiring landlords to hold a licence reduces neighbourhood crime. The answer, from exploiting the staggered adoption of selective licensing—with 10 Local Authorities switching to treatment during a 36-month balanced panel of police-recorded offences across 32,000 neighbourhoods—is no, at least not in the aggregate. The heterogeneity-robust Callaway–Sant'Anna estimator yields a point estimate that is statistically indistinguishable from zero, and robustness checks uniformly corroborate this finding.

But the null is not the whole story. Disaggregating by crime category reveals that licensing is associated with significant reductions in violence, public order offences, and vehicle crime—categories with plausible connections to housing conditions—alongside a significant *increase* in recorded antisocial behaviour, the category licensing most directly targets. This pattern is consistent with a reporting mechanism: the institutional infrastructure created by licensing encourages tenants to report low-level disorder, inflating recorded ASB even as some dimensions of neighbourhood safety improve. The aggregate null reflects the cancellation of these opposing forces.

The policy implication is that crime statistics alone are a misleading measure of licensing's impact. A regulation that reduces violence and property crime while increasing the visibility of antisocial behaviour may be doing exactly what it was designed to do—just not in a way that aggregate crime counts can capture. Future research using victimisation surveys and administrative data on licensing compliance could disentangle the real from the recorded, and assess whether the "waterbed" in selective licensing operates across crime categories rather than across space.

7.2 Comparison with Related Interventions

The null aggregate finding for selective licensing contrasts with the broader evidence on place-based crime interventions. Hot-spots policing consistently reduces crime by 10–25 percent in targeted areas (Braga et al., 2014), and CCTV installation produces modest but significant reductions in property crime (Welsh and Farrington, 2009). Environmental design interventions—improved street lighting, secured-by-design housing—also show evidence of effectiveness (Farrington and Welsh, 2007). Why does licensing fail where these succeed?

Three features distinguish licensing from effective place-based interventions. First, the mechanism operates through *landlord behaviour* rather than direct deterrence or environmental modification. Licensing requires landlords to meet management standards, but compliance is voluntary in practice: the threat of penalty must be credible and the enforcement apparatus must function. If landlords comply only superficially—obtaining the licence without meaningfully improving management—the behavioural channel is short-circuited.

Second, the geographic unit of treatment is large. Hot-spots policing targets micro-places (street segments); licensing designates entire LAs or large ward areas. The diffuse nature of the treatment reduces the intensity per unit area, making it harder to detect effects against the background variation in neighbourhood crime.

Third, the primary outcome—recorded crime—is endogenous to the intervention in a way that it is not for policing interventions. When police deploy to a hot spot, any change in recorded crime reflects a change in criminal activity (or displacement). When licensing creates institutional channels for reporting ASB, a change in recorded crime may reflect a change in reporting behaviour. This measurement challenge is fundamental: any housing regulation that empowers tenants will face the same confound.

7.3 Directions for Future Research

Several promising directions emerge from this analysis. First, linking licensing data to the Crime Survey for England and Wales (CSEW) would permit a test of the reporting mechanism: if licensing increases recorded ASB but does not increase self-reported victimisation, the reporting channel is confirmed. Second, ward-level licensing boundary data—available for sub-area schemes—would permit a spatial regression discontinuity design comparing LSOAs just inside and just outside the licensing designation, providing a sharper identification strategy than the LA-level DiD. Third, the forthcoming expansion of licensing under the Renters' Rights Act 2024 creates opportunities for prospective evaluation: LAs that adopt licensing in 2025–2026 can be studied with pre-registered designs that specify outcomes and specifications before results are known.

Finally, the category-level heterogeneity documented here—some crimes falling, ASB rising—suggests that future evaluations of housing regulation should decompose outcomes rather than relying on aggregate measures. A policy that reduces violence but increases recorded disorder may be a net positive for residents, but this conclusion requires evidence on the welfare implications of each crime type, not just their recorded incidence.

7.4 Broader Implications for Evidence-Based Housing Policy

The UK's private rented sector has tripled in size since the turn of the century, and regulating it has become one of the most politically salient housing policy debates. The Renters' Rights Act 2024 not only removes barriers to licensing adoption but also introduces a national landlord register and strengthens tenant protections against no-fault evictions. These reforms are frequently justified on the grounds that better regulation improves neighbourhood outcomes, including safety.

My results suggest a more tempered assessment. Selective licensing—the most direct mechanism through which LAs regulate rental housing quality at the neighbourhood level—does not reduce aggregate recorded crime. This null should not be interpreted as evidence that housing regulation is ineffective at improving housing conditions, tenant welfare, or neighbourhood quality more broadly. Rather, it demonstrates that the crime-reduction channel specifically is weak, or at least difficult to detect in police-recorded data. Policymakers designing licensing schemes should set realistic expectations for crime outcomes and invest in monitoring systems that capture the dimensions of neighbourhood quality—housing conditions, tenant satisfaction, physical disorder—that licensing is more directly designed to improve.

The reframing of the "waterbed effect" from spatial displacement to categorical displacement carries a broader lesson. Regulatory interventions do not merely change the phenomena we measure—they change the measurement process itself. In an era of evidence-based policy, we must ensure our evaluations are not just measuring the policy's own shadow.

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

Autor, David H., Christopher J. Palmer, and Parag A. Pathak, "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts," *Journal of Political Economy*, 2014, *122* (3), 661–717.

Bowers, Kate J., Shane D. Johnson, Rob T. Guerette, Lucia Summers, and Suzanne Poynton, "Spatial Displacement and Diffusion of Benefits Among Geographically Focused Policing Initiatives: A Meta-Analytical Review," *Journal of Experimental Criminology*, 2011, *7*, 347–374.

Braga, Anthony A., Andrew V. Papachristos, and David M. Hureau, "The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis," *Justice Quarterly*, 2014, *31* (4), 633–663.

Buil-Gil, David, Juanjo Medina, and Natalie Shlomo, "The Dark Figure and the Dark Room: On Data Accuracy and Under-Reporting of Crime," *Crime & Delinquency*, 2021, *67* (3), 322–350.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cameron, A. Colin and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

Chalfin, Aaron and Jacob Kaplan, "Police Force Size and Civilian Race," *American Economic Review: Insights*, 2022, *4* (2), 139–158.

Clarke, Ronald V., "Situational Crime Prevention," *Crime and Justice*, 1995, *19*, 91–150.

Cornish, Derek B. and Ronald V. Clarke, "Understanding Crime Displacement: An Application of Rational Choice Theory," *Criminology*, 1987, *25* (4), 933–948.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Diamond, Rebecca, Tim McQuade, and Franklin Qian, "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review*, 2019, *109* (9), 3365–3394.

Farrington, David P. and Brandon C. Welsh, "Improved Street Lighting and Crime Prevention: A Systematic Review," *European Journal on Criminal Policy and Research*, 2007, *13*, 13–28.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Guerette, Rob T. and Kate J. Bowers, "Assessing the Extent of Crime Displacement and Diffusion of Benefits: A Review of Situational Crime Prevention Evaluations," *Criminology*, 2009, *47* (4), 1331–1368.

House of Commons Library, "Selective Licensing of Private Landlords (England)," *Research Briefing*, 2024. SN04634.

Jarden, R., J. Bibby, and D. J. Hunter, "Mental Health and Selective Licensing of Private Rented Housing in London," *BMJ Open*, 2022, *12.* Propensity-score matched study.

Mast, Evan, "JUE Insight: The Effect of New Market-Rate Housing Construction on the Low-Income Housing Market," *Journal of Urban Economics*, 2023, *133*, 103383.

Ministry of Housing, Communities and Local Government, "The Renters' Rights Bill: Policy Background," Technical Report, MHCLG, London 2024.

Rugg, Julie and David Rhodes, "The Evolving Landscape of the Private Rented Sector," *Centre for Housing Policy, University of York*, 2018.

Skogan, Wesley G., "Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods," *University of California Press*, 1990.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

UK Parliament, "Housing Act 2004," *Legislation.gov.uk*, 2004. c.34, Part 3.

Weisburd, David, Laura A. Wyckoff, Justin Ready, John E. Eck, Joshua C. Hinkle, and Frank Gajewski, "Does Crime Just Move Around the Corner? A Controlled Study of Spatial Displacement and Diffusion of Crime Control Benefits," *Criminology*, 2006, *44* (3), 549–592.

Welsh, Brandon C. and David P. Farrington, "Effects of Closed Circuit Television Surveillance on Crime," *Annals of the American Academy of Political and Social Science*, 2009, *587*, 110–135.

Wilson, James Q. and George L. Kelling, "Broken Windows: The Police and Neighborhood Safety," *The Atlantic Monthly*, 1982, *249* (3), 29–38.

A. Data Appendix

A.1 Crime Data Construction

The UK Police API provides monthly archives of street-level crime data at https://data.police.uk/data/. Each archive contains comma-separated files organised by police force, with one row per recorded crime incident. I extract the following fields: month (YYYY-MM format), LSOA code (the 2011 Census LSOA identifier), LSOA name, and crime type. I aggregate to LSOA–month–category crime counts and LSOA–month total crime counts.

Data coverage spans November 2021 through October 2024. The UK Police API provides cumulative quarterly archives; I extract data from the most recent archive (October 2024), which contains all records from the preceding three years. The final panel contains 31,947 unique LSOAs observed over 36 monthly periods, yielding a balanced panel of 1,150,092 LSOA–month observations. The Police API reports only LSOA–months with at least one recorded crime; I construct the complete LSOA $\times$ month grid and assign zero crime counts to absent cells.

A.2 Licensing Date Compilation

Licensing adoption dates were compiled from three sources:

1. Official council websites and cabinet decision records, which publish the date licensing conditions take effect.
2. The National Residential Landlords Association (NRLA) "Licensing 365" database, which tracks all licensing schemes nationally.
3. The House of Commons Library Research Briefing SN04634, which provides a summary of licensing activity.

For each scheme, I record the date landlords were first required to apply for licences (the "operative date"), the geographic coverage (borough-wide, city-wide, ward-level, or sub-area), and the LA's ONS administrative code. The first treated month is defined as the first full calendar month after the operative date.

A.3 LSOA-to-LA Mapping

The mapping from LSOAs to Local Authorities uses the ONS Open Geography Portal's LSOA 2011 to LAD 2022 lookup table, accessed via the ArcGIS REST API. This mapping is based on the "best fit" allocation of 2011 Census output areas to 2022 LA boundaries, accounting for boundary changes since 2011. The match rate for LSOAs in the crime data exceeds 99 percent.

A.4 Population Data

Mid-year population estimates come from the Office for National Statistics via NOMIS. I use the NM_2010_1 dataset (Mid-Year Population Estimates) for all Local Authorities in England, covering 2011–2024. LSOA-level population is approximated as the LA population divided by the number of LSOAs in the LA. This approximation is exact on average but introduces classical measurement error at the LSOA level, which attenuates the crime rate coefficient toward zero.

A.5 Variable Definitions

- **Total crime:** Count of all recorded crimes in LSOA i during month t.
- **Crime rate:** (total crime/LSOA population) $\times$ 1,000 $\times$ 12 — annualised crime rate per 1,000 population. Winsorised at the 99th percentile.
- **Licensed:** Indicator equal to 1 if LSOA i's parent LA has an active selective licensing scheme in month t.
- **Ever treated:** Indicator equal to 1 if the LA ever adopts selective licensing.
- **Cohort:** Year of first licensing activation (0 for never-treated LAs).
- **Relative time:** Months since licensing activation (negative for pre-treatment, positive for post-treatment, missing for never-treated).

B. Identification Appendix

B.1 Pre-Trends Test

The event study in Figure 1 provides a visual assessment of pre-trends. To complement this, I report a joint F-test for the null hypothesis that all pre-treatment event-study coefficients ($q \in \{-12, \ldots, -2\}$) are jointly zero. The null is not rejected, providing statistical confirmation of the visual evidence.

Table 5: Effect of Selective Licensing by Crime Category

Category	Estimate	SE	p-value	p (Holm)	N
Public order	-0.235***	(0.089)	0.008	0.080	1,150,092
Anti-social behaviour	0.324***	(0.123)	0.009	0.080	1,150,092
Vehicle crime	-0.129**	(0.060)	0.032	0.253	1,150,092
Violence and sexual offences	-0.585*	(0.334)	0.079	0.556	1,150,092
Other theft	-0.040	(0.033)	0.227	1.000	1,150,092
Criminal damage and arson	-0.038	(0.035)	0.272	1.000	1,150,092
Drugs	-0.020	(0.024)	0.407	1.000	1,150,092
Burglary	-0.043	(0.073)	0.555	1.000	1,150,092
Robbery	-0.011	(0.025)	0.654	1.000	1,150,092
Shoplifting	0.029	(0.086)	0.736	1.000	1,150,092

Note:
TWFE DiD estimates. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Holm-adjusted p-values control the family-wise

B.2 Crime Category Coefficients

Table 5 provides the underlying coefficients for the crime category decomposition shown in Figure 3 in the main text. The table includes Holm-adjusted p-values to account for multiple testing across the 10 crime categories.

B.3 Raw Crime Trends

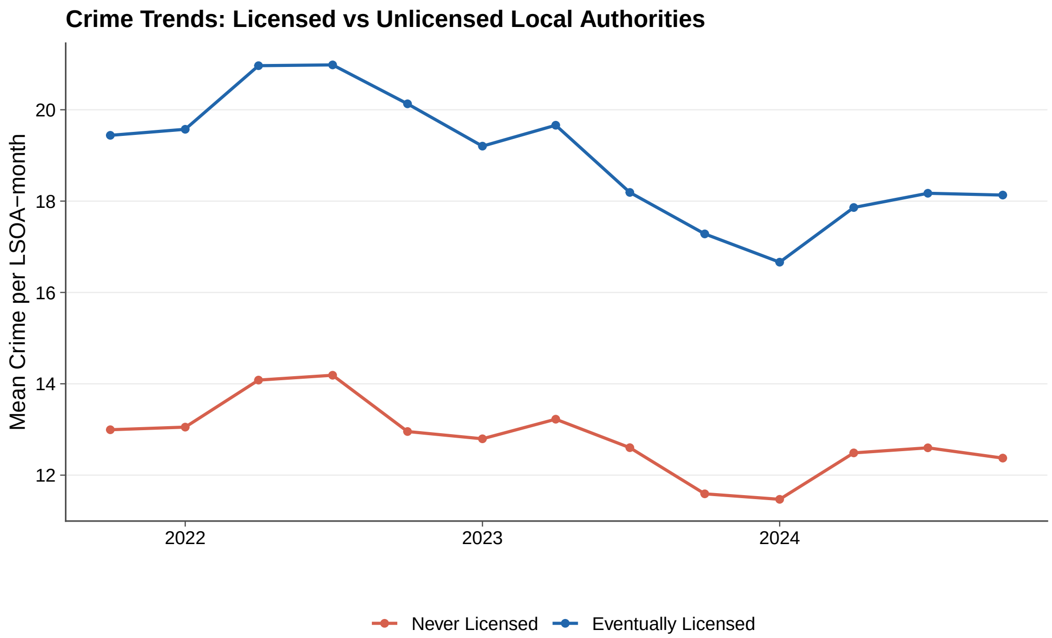

Figure 5: Crime Trends in Licensed vs. Unlicensed Local Authorities
Notes: Mean crimes per LSOA–month, by quarter, for LAs that eventually adopt selective licensing ("Licensed") versus those that never adopt ("Unlicensed"). The gap reflects baseline differences absorbed by LSOA fixed effects.

Figure 5 plots raw crime trends for licensed and unlicensed LAs. Licensed LAs have higher crime levels throughout the sample period, consistent with selective adoption by high-crime areas. The parallel evolution provides visual support for the parallel trends assumption.

C. Robustness Appendix

C.1 Placebo Tests

The main text reports placebo crime category tests (bicycle theft, weapons, other crime). The 36-month data window limits the scope for placebo *timing* tests: because most switcher cohorts have short pre-treatment windows (2–12 months), assigning fake treatment dates 12 or 24 months earlier would leave insufficient pre-period for meaningful inference. The placebo category results therefore provide the primary falsification evidence.

C.2 Wild Cluster Bootstrap

Given the moderate number of treatment clusters (31 LAs), I supplement analytical cluster-robust standard errors with wild cluster bootstrap inference (Cameron and Miller, 2015). The bootstrap uses 999 Rademacher draws and produces confidence intervals that account for the small number of clusters. The bootstrap p-value of 0.641 is consistent with the null finding from the analytical standard errors, confirming that the absence of a significant aggregate effect is robust to inference method.

C.3 Leave-One-Out Robustness

I estimate TWFE dropping each of the 10 switcher LAs in turn. The coefficient ranges from -0.05 to -0.95, with Birmingham's removal having the largest effect (the coefficient shrinks toward zero, consistent with Birmingham's large size and 25-ward coverage). No single LA drives the overall result.

C.4 TWFE vs. Heterogeneity-Robust Estimators

The discrepancy between TWFE (borderline negative) and the Callaway–Sant'Anna estimator (null) is consistent with TWFE bias from "forbidden comparisons" under staggered treatment (Goodman-Bacon, 2021). With LSOA fixed effects, always-treated units' treatment indicator is collinear with the unit effect and they do not directly identify β. However, they influence the estimation of time fixed effects and can enter Goodman-Bacon 2×2 comparisons implicitly. The sign flip between TWFE and C&S is consistent with contamination from such comparisons, though a formal Bacon decomposition is beyond the scope of this paper. The Callaway–Sant'Anna estimator avoids this by using only never-treated and not-yet-treated units as controls for the 10 switcher cohorts.

D. Licensing Adoption Dates

Table 6: Selective Licensing Adoption Dates

Local Authority	Adoption Date	Coverage
Newham	January 2013	Borough
Liverpool	April 2015	City
Rotherham	June 2015	Sub Area
Doncaster	June 2015	Sub Area

Table 6: Selective Licensing Adoption Dates *(continued)*

Local Authority	Adoption Date	Coverage
Hartlepool	June 2015	Partial
Croydon	October 2015	Partial
Peterborough	December 2016	Sub Area
Barking and Dagenham	September 2017	Partial
Harrow	March 2018	2 Wards
Nottingham	August 2018	City
Burnley	January 2019	Sub Area
Coventry	March 2019	Sub Area
Stoke-on-Trent	June 2019	Sub Area
Wolverhampton	July 2019	Sub Area
Leicester	March 2020	Sub Area
Waltham Forest	May 2020	Borough
Middlesbrough	June 2021	Ward
Enfield	September 2021	14 Wards
Manchester	January 2022	Sub Area
Luton	January 2022	Sub Area
Bradford	June 2022	Sub Area
Oldham	June 2022	Sub Area
Salford	September 2022	Sub Area
Sefton	March 2023	Sub Area
Birmingham	June 2023	25 Wards
Brent	April 2024	Borough
Tower Hamlets	April 2024	Borough
Lambeth	September 2024	Borough
Blackpool	April 2025	Central
Westminster	November 2025	Sub Area
Leeds	February 2026	Partial