stats

MDPI

Article

Correlation-Adjusted Elastic-Net Penalties for Neighborhood Crime Deprivation Modeling in England

Taiwo Marcus Akinmuyisitan [1,†], Olayiwola Babarinsa [2,*], Boluwaji Bernard Akinmuyisitan [3], John Cosmas [1] and Temitope Adegbeyeni [4]

[1] College of Engineering, Brunel University of London, Kingston Lane, Uxbridge UB8 3PH, UK; takinmuyisitan@globalbanking.ac.uk (T.M.A.); john.cosmas@brunel.ac.uk (J.C.)
[2] Department of Mathematics, Federal University Lokoja, Lokoja 260102, P.M.B 1154, Kogi State, Nigeria
[3] School of Business Studies, University of Greater Manchester, Deane Road, Greater Manchester, Bolton BL3 5AB, UK; ba1res@bolton.ac.uk
[4] School of Science, Engineering and Environment, University of Salford, Salford M5 4WT, UK; t.o.adegbeyeni@edu.salford.ac.uk
* Correspondence: olayiwola.babarinsa@fulokoja.edu.ng
† Current address: Global Banking School, Computing Department, GEDU House-260 High Street, Stratford E15 2JA, UK.

Abstract

A collection of interrelated social variables determines crime deprivation at the neighborhood level in England. In penalized regression, strong intercorrelations pose a fundamental problem for variable selection. Previous research has used Elastic-Net (ENET) or the Least Absolute Shrinkage and Selection Operator (LASSO) to analyze individual waves of the English Indices of Deprivation (IoD) at the Lower Super Output Area (LSOA) level. This leaves questions about what the recently released IoD 2025 reveals about the crime–deprivation relationship, whether variable selection is stable across IoD waves, and whether the Correlation-Adjusted Elastic-Net (CAEN), which embeds actual pairwise Pearson correlations into the penalty matrix, achieves greater sparsity than LASSO and ENET. Using a single consistent Python (3.14.2) implementation that removes cross-software confounds observed in previous CAEN1 assessments, this work applies LASSO, ENET, and CAEN1 to all four publicly available releases of the IoD: 2010 (n = 32,482), 2015 ($n = 32,844$), 2019 ($n = 32,844$), and 2025 ($n = 33,755$). In the CAEN penalty, the diagonal elements guarantee positive semi-definiteness and global convergence of coordinate descent. To ensure an equitable comparison with the classic ENET, the CAEN penalty rescales the penalty factor to correct for double shrinkage. In three out of four waves, CAEN1 outperforms LASSO in terms of sparsity over 50 stratified 70/30 splits using 10-fold cross-validation (CV). Building a six-predictor model that neither LASSO nor ENET could, the CAEN1 set both Income and Employment to zero in 2015. For the mean square error, the differences in prediction accuracy are negligible in magnitude (ΔMSE < 0.002). IDACI, Health and Disability, and Living Environment are the most consistently prominent determinants of neighborhood crime deprivation, according to a 15-year longitudinal study. A corrected resampled t-test confirms that the methods differ little inout-of-sample MSE in three of the four waves, with only a negligible CAEN1 excess reaching significance in 2010. A simulation further shows that in higher-dimensional, strongly collinear, and low signal-to-noise settings, CAEN2 coupling recovers correlated predictor groups that LASSO fragments.

Check for updates

Academic Editor: B. M. Golam Kibria

Received: 24 July 2026
Revised: 10 September 2026
Accepted: 17 September 2026
Published: 28 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: crime deprivation; regularized regression; variable selection; indices of deprivation; longitudinal stability

1. Introduction

The link between social deprivation and crime has been the focus of criminology, public Health, and social policy for over a century. High-Income deprivation, unemployment, low educational attainment, poor Health, and subpar living conditions are all associated with greater rates of criminal behavior at the neighborhood level; see [1–3]. Based on social disorganization theory, communities experience concentrated disadvantage because citizens have diminished capacity and willingness to address neighborhood disorder, fostering an environment conducive to crime [4–6]. According to routine activity theory, in the absence of capable guardians, crime opportunities arise when potential offenders and suitable targets come together [7].

Hirschfield and Bowers [8] showed that a combination of inadequate housing, educational disadvantage, and Income deprivation best described concentrations of property crime in Merseyside. Hipp [9] found that concentrated poverty and home instability are substantially correlated with neighborhood-level crime rates in several US cities, with social control and cohesion mediating this relationship. Deller and Deller [10] studied rural US crime patterns utilizing social capital, transactional economic indicators, and U.S. county datasets. Steenbeek and Weisburd [11] demonstrated the temporal stability of crime hot spots in which prior crime levels predict future crime concentrations at micro-places.

Penalized regression is a statistical and machine learning technique that involves combining a penalty term to a model's residual sum of squares or loss function. Traditional penalized regression methods, such as Elastic-Net (ENET), Ridge, and Least Absolute Shrinkage and Selection Operator (LASSO), have drawbacks when dealing with correlated features [12,13]. Penalized regression techniques, however, handle associated predictors differently. In the mid-2010s, penalized regression techniques in criminology began to grow in use. Berk [14] applied penalized regression to criminal justice work to sift through large intake datasets. This predicts inmate misconduct and recidivism risk while sidestepping the overfitting problems that standard logistic regression often encounters. Caplan et al. [15] brought Elastic-Net into spatial policing tools. This allows analysts to feed dozens of potential environmental crime generators into one model, with the penalty term dropping irrelevant variables on its own and handling the heavy multicollinearity that urban data tends to have. - Perry tested penalized regression models against kernel density estimation, yielding sparse risk maps that meaningfully reduced false-positive patrol assignments [12]. On small-to-moderate biomedical datasets, Babarinsa et al. [16] showed that CAEN performs better than EN.

This study compares LASSO, ENET, and CAEN1 across the full suite of four IoD waves (2010, 2015, 2019, and 2025). Each Lower Layer Super Output Area (LSOA) in the English IoD contains about 1500 residents. The first version to use 2021 Census LSOAs is the 2025 IoD [17]. The IoD crime domain combines rates for four recorded crime types (violence, burglary, theft, and criminal damage) and standardizes them to facilitate cross-area comparison. Recorded crime rates are used to create the IoD crime domain, and each wave's cross-sectional design finds predictive relationships. As a result, neighboring LSOAs have similar unmeasured features, and the penalized regression models regard LSOAs as independent components.

Section 2 concisely describes LASSO, ENET, CAEN, and their implementations. Section 3 presents the four IoD waves, their predictor correlation structures, and the experimen-

tal protocol. Section 4 reports results on prediction accuracy, model sparsity, and variable selection across all four waves. Section 5 discusses the findings, including CAEN1's sparsity advantage and the structural shift observed in the 2025 IoD release.

2. Statistical Methods

Let $Y \in \mathbb{R}^n$ be a vector of n continuous responses and $X \in \mathbb{R}^{n\times p}$ a matrix of p predictors observed on the same n units [18]. The linear regression model is

$$Y = X\beta + \epsilon,\ \varepsilon \sim N\left(0,\ \sigma^2 I\right) \tag{1}$$

where Y is the $n \times 1$ response vector, X the $n \times p$ design matrix, and $\beta = (\beta_1, \ldots, \beta_p)^T$ the $p \times 1$ coefficient vector. Errors are assumed identically and independently distributed (IID) with zero mean and finite variance σ^2. This IID assumption is assessed empirically via Moran's I [19].

When predictors are highly correlated, ordinary least squares estimates are unstable, and selecting relevant predictors is unreliable. Thus, penalized regression addresses this by adding a penalty term to the least-squares objective. The LASSO, ENET, and CAEN handle some of the problems associated with selection.

The LASSO shrinks coefficients toward zero and sets small ones exactly to zero to achieve variable selection [20]. This is given as

$$\hat{\beta}_{lasso} = arg\ \min_{\beta}\left(\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1^1\right), \tag{2}$$

where $\|\beta\|_1^1 = \sum_{j=1}^{p}|\beta_j|$ is the l_1-norm of β and $arg\min_{\beta}$ is the argument of the minimum [21].

LASSO's main limitation with grouped, correlated predictors is that it selects at most one variable from each group. This reduces interpretability when the substantive distinctions among variables in a group matter. The ENET adds an L_2 penalty to encourage grouped selection as defined below [22,23]:

$$\hat{\beta}_{enet} = arg\ \min_{\beta}\left(\|Y - X\beta\|_2^2 + \lambda_1\|\beta\|_1^1 + \lambda_2\|\beta\|_2^2\right). \tag{3}$$

Nonetheless, the L_2 component does not explicitly use the pairwise correlation structure among the predictors. Tan [24] extended ENET with the CAEN, which replaces the L_2 identity penalty with a matrix W whose off-diagonal entries encode absolute pairwise Pearson correlations. The CAEN incorporates pairwise correlations between adjacent predictors or all pairs of predictors directly into a quadratic penalty term. Thus,

$$\hat{\beta}_{CAEN} = \operatorname{argmin}_{\beta}\left(\|Y - X\beta\|_2^2 + \lambda_1\|\beta\|_1^1 + \lambda_2\beta^T W\beta\ \right), \tag{4}$$

where W is a positive semi-definite matrix with strictly positive diagonal elements such that $W_{jj} > 0$. The quadratic form $\beta^T W\beta$ is non-negative for all β and β^T is the transpose of the vector of regression coefficients. The correlation-coupling penalty is denoted as $P(\beta) = \lambda_2\beta^T W\beta$, so that the objective in Equation (4) can be written as $\|Y - X\beta\|^2 + \lambda_1\|\beta\|_1 + P(\beta)$. Thus,

$$P_{CAEN1}(\beta) = \sum_{j=1}^{p-1}\left(\beta_j - r_{j,j+1}\beta_{j+1}\right)^2 + \beta_p^2$$

$$P_{CAEN2}(\beta) = \sum_{j=1}^{p-1}\sum_{k>j}^{p}\left(\beta_j - r_{j,k}\beta_k\right)^2 + \beta_p^2$$

where $r_{j,k}$ denotes the sample Pearson correlation between predictors X_j and X_k. CAEN2 extends this to all pairwise correlations such that W_{jj} is strictly convex [24]. As a result, highly correlated predictors yield a higher joint penalty, making the technique correlation-aware in contrast to ENET. However, when predictors are highly correlated, a predictor may receive a negative LASSO coefficient even though it has a positive marginal correlation with the outcome. This suppressor effect arises when the predictor explains variance in another predictor beyond its own direct association with the outcome [25]. Table 1 details the W matrix for CAEN1 and CAEN2.

Table 1. Structure of the CAEN penalty matrix W for CAEN1 and CAEN2.

Metric	CAEN1	CAEN2
Matrix structure	Tridiagonal	Dense (all pairs connected)
Off-diagonal $W_{j,k}$	$-r_{j,k} \cdot \mathbb{I}(\lvert j-k \rvert = 1)$	$-r_{j,k} \forall j \neq k$
Diagonal $W_{j,j} (j < p)$	$1 + r^2_{j-1,j} \cdot \mathbb{I}(j > 1)$	$(p-j) + \sum_{i<j} r^2_{i,j}$
Diagonal $W_{p,p}$	$1 + r^2_{p-1,p}$	$1 + \sum_{i=1}^{p-1} r^2_{i,p}$

LASSO and CAEN applied zero coefficients through a soft-thresholding rule during coordinate descent, but the threshold differs between the two methods. For LASSO, the coordinate-descent update for each predictor j is [26]

$$\hat{\beta}_j = sign(z_j) \cdot max(|z_j| - \lambda, 0), \tag{5}$$

where

$$\text{sign}(z_j) = \begin{cases} +1 & \text{if } z_j > 0 \\ -1 & \text{if } z_j < 0 \\ 0 & \text{if } z_j = 0 \end{cases} \tag{6}$$

where z_j is the partial residual correlation of predictor j with the outcome after removing the contributions of all other predictors. The soft threshold, $max(|z_j| - \lambda, 0)$, returns exactly zero when $|z_j| \leq \lambda$. The predictor is excluded from the model because its residual signal is too weak relative to the penalty. The threshold λ is the same for all predictors, so LASSO applies an equal zeroing standard regardless of how correlated a predictor is with its neighbors.

CAEN modifies the partial residual correlation by augmenting the normal equations with the correlation matrix W. The coordinate-descent update becomes:

$$\hat{\beta}_j = sign(r_j) \cdot \frac{max(|r_j| - \lambda, 0)}{A_{jj}}, \tag{7}$$

where r_j is the partial residual, and A_{jj} is of the form

$$A_{jj} = X_j^T X_j + \lambda_2 \left(1 + \sum_{k \neq j} r^2_{j,k}\right) \tag{8}$$

The $max(|r_j| - \lambda, 0)$ follows Equation (6). A_{jj} is larger for predictors that are highly correlated with others, raising the effective zeroing threshold for those predictors. Equation (8) ensures that predictors with high aggregate squared correlations with their neighbors receive a larger A_{jj}. This raises the effective zeroing threshold that coordinate descent drives to zero.

One documented flaw of regularized estimators, including L_1 and quadratic penalties, is double-shrinkage bias [22]. Hence, as both penalty components push the coefficient estimates toward zero, the resulting estimators are more severely attenuated. To address this, they suggested rescaling the naïve ENET coefficient vector by $(1 + \lambda^2)$ to mitigate this bias. The current implementation uses this rescaling factor $(1 + \lambda^2)$ to adjust the naive CAEN coefficient vector following convergence. This rescaling factor provides an adjusted/corrected CAEN estimator consistent with Babarinsa. This ensures an equitable comparison with ENET fitted by scikit-learn's ElasticNetCV, which also applies this correction by default.

3. Data and Experimental Design

In this section, we obtain the dataset from the Ministry of Housing, Communities, and Local Government database and evaluate predictor correlations using the dataset to train LASSO, ENET, and CAEN.

3.1. English Indices of Deprivation

In England, the Ministry of Housing, Communities and Local Government (MHCLG) operationalizes the English IoD at the Lower Layer Super Output Area (LSOA) level, with each LSOA capturing approximately 1500 residents [27,28]. The IoD crime domain combines rates for four recorded crime types (violence, burglary, theft, and criminal damage) and standardizes them to facilitate cross-area comparison [17]. IoD 2010 ($n = 32,482$), IoD 2015 ($n = 32,844$), IoD 2019 ($n = 32,844$), and IoD 2025 ($n = 33,755$) are the four IoD periods covered in this research, with crime score as the response variable. The 2025 IoD increase reflects the 2021 Census boundary revision, which produced more homogeneous LSOAs. The eight predictors of the IoD, ordered so that the related Income-based domains are adjacent, are Employment, Education, skills and training, Health and Disability, housing and service Barriers, Living Environment, Income, income deprivation impacting children index (IDACI), and income deprivation affecting older people index (IDAOPI). To prevent circularity, the overall index of multiple deprivation (IMD) score is not included. The modeling pipeline applies standardization, and all scores are continuous, as supplied by MHCLG.

3.2. Experimental Protocol

Using a 10-fold CV, each dataset is divided 50 times (random state = i for replication i = 0, . . ., 49) into a 70% training set and a 30% held-out test set [29]. We use only training observations to normalize predictors to zero mean and unit variance. Test observations undergo the same modification without being refitted. As explained in Section 2, nested cross-validation is used for all hyperparameter selection inside the training set. The chosen l_1–ratio grid {0.50, 0.70, 0.90, 0.95, 0.99} adheres to scikit-learn's suggested defaults. LASSO and ENET are well known in scikit-learn's LassoCV and ElasticNetCV [30], while quadrupen R was formally used for CAEN1 [31]. Now, CAEN1 uses a custom – NumPy 2.5.0 (Numerical Python) coordinate-descent solver in the same environment and implements Equation (7) directly. Using a single language eliminates cross-software confounds in intercept centering, convergence tolerance, and lambda-grid creation. Values in the $l1_{ratio}$ grid that are less than 0.50 are eliminated because they approach Ridge regression and are unable to generate sparse solutions. Similarly, $\rho = 1.0$ is excluded to keep ENET distinct from LASSO during the search. When predictors are highly correlated, a finer spacing around $\rho = 1.0$ provides greater resolution in the region where ENET is most likely to deviate from LASSO.

Consider that the 50 repeated 70/30 splits reuse the same observations and are therefore not independent. We assess pairwise differences in out-of-sample MSE between

methods using the corrected resampled t-test of Nadeau and Bengio [32]. This adjusts the variance of the mean difference by the factor $(\frac{1}{J} + \frac{n_{test}}{n_{train}})$ to account for the dependence between splits (J = 50). For each of the 50 splits, we choose hyperparameters by 10-fold inner cross-validation on the 70% training set only. We never use the 30% test set for tuning. For LASSO and ENET, scikit-learn's LassoCV and ElasticNetCV construct the regularization path automatically over 100 λ values. This is spaced logarithmically from λ_{max} down to $\frac{\lambda_{max}}{1000,}$ with ENET searching the l_1–ratio grid {0.50, 0.70, 0.90, 0.95, 0.99}. For CAEN1, the custom coordinate-descent algorithm is parameterized with λ_1 values of {0.001, 0.003, 0.01, 0.03, 0.1, 0.3}. Moreover, the training set response standard deviation and the λ_2 values of {0.01, 0.05, 0.1, 0.3} are used, where the combination $(\lambda_1,\ \lambda_2)$ resulting in the minimum inner-CV MSE is used. Coordinate descent stops when the change in coefficients falls to 1×10^{-7} (a maximum of 500 sweeps allowed). The same standardization, computed on the training fold only, is applied to the validation and test folds without refitting.

This study adopts CAEN1 (adjacent-predictor coupling) rather than CAEN2 (all-pair coupling) for specific reasons. First, the eight IoD domains form a low-dimensional, ordered set of deprivation indicators, with the strongest dependencies between neighboring domains. The tridiagonal CAEN1 matrix captures this local correlation structure, whereas the dense CAEN2 matrix couples only weakly related domain pairs and adds little beyond adjacent coupling. Furthermore, CAEN1 is markedly more parsimonious and numerically stable. Its penalty matrix has $O(p)$ non-zero entries and remains well-conditioned. In contrast, CAEN2 introduces $O(p^2)$ off-diagonal couplings, whose accumulated correlation weighting can inflate the diagonal and over-penalize coefficients, making the coordinate-descent solution more sensitive to sampling noise. Finally, CAEN1 keeps the comparison with LASSO and ENET interpretable, since only a coefficient's immediate neighbors raise its zeroing threshold. This allows the suppressor and grouping effects reported in Section 4 to be attributed to identifiable predictor pairs rather than to a diffuse all-pairs penalty. CAEN2 is retained in Section 2 as the general formulation and is a natural direction for higher-dimensional applications, where all-pair coupling is expected to matter more.

3.3. Predictor Correlation Structure

For each of the four IoD waves, we account for the predictor correlation structure. For each wave in Tables 2–5, we build these structures from pairwise Pearson correlations calculated from the IoD domain scores.

Table 2. Pairwise Pearson correlations for IoD 2010 ($n = 32,482$).

	Income	Employment	Education	Health & Dis.	Barriers	Living Env.	IDACI	IDAOPI
Income	1.000	0.900	0.786	0.787	0.040	0.478	0.947	0.890
Employment	0.900	1.000	0.782	0.857	−0.109	0.391	0.802	0.770
Education	0.786	0.782	1.000	0.685	−0.168	0.262	0.725	0.647
Health & Dis.	0.787	0.857	0.685	1.000	−0.125	0.430	0.724	0.761
Barriers	0.040	−0.109	−0.168	−0.125	1.000	0.202	0.112	0.065
Living Env.	0.478	0.391	0.262	0.430	0.202	1.000	0.487	0.555
IDACI	0.947	0.802	0.725	0.724	0.112	0.487	1.000	0.828
IDAOPI	0.890	0.770	0.647	0.761	0.065	0.555	0.828	1.000

Pairwise Pearson correlations calculated from the IoD domain scores for each wave are shown in Tables 2–5. Throughout 2010–2025, the primary deprivation cluster (Income, Employment, Education, Health, IDACI, IDAOPI) maintains pairwise correlations of r = 0.62–0.95. In contrast, Barriers and Living Environment continue to be structurally distinct from the main cluster with $|r| < 0.50$ in every wave. Variable-selection instability for LASSO in 2010 and 2015 is mostly caused by cells with $|r| \geq 0.80$, which are concentrated in the Income–Employment–IDACI–IDAOPI quadrant.

Table 3. Pairwise Pearson correlations for IoD 2015 ($n = 32,844$).

	Income	Employment	Education	Health & Dis.	Barriers	Living Env.	IDACI	IDAOPI
Income	1.000	0.953	0.836	0.814	0.157	0.255	0.944	0.861
Employment	0.953	1.000	0.834	0.855	0.053	0.196	0.865	0.770
Education	0.836	0.834	1.000	0.713	0.047	0.106	0.795	0.648
Health & Dis.	0.814	0.855	0.713	1.000	−0.037	0.260	0.759	0.746
Barriers	0.157	0.053	0.047	−0.037	1.000	0.254	0.182	0.236
Living Env.	0.255	0.196	0.106	0.260	0.254	1.000	0.269	0.388
IDACI	0.944	0.865	0.795	0.759	0.182	0.269	1.000	0.792
IDAOPI	0.861	0.770	0.648	0.746	0.236	0.388	0.792	1.000

Table 4. Pairwise Pearson correlations for IoD 2019 ($n = 32,844$).

	Income	Employment	Education	Health & Dis.	Barriers	Living Env.	IDACI	IDAOPI
Income	1.000	0.954	0.833	0.799	0.044	0.191	0.940	0.847
Employment	0.954	1.000	0.825	0.841	−0.056	0.124	0.872	0.737
Education	0.833	0.825	1.000	0.721	−0.064	0.053	0.809	0.624
Health & Dis.	0.799	0.841	0.721	1.000	−0.158	0.167	0.754	0.683
Barriers	0.044	−0.056	−0.064	−0.158	1.000	0.214	0.061	0.162
Living Env.	0.191	0.124	0.053	0.167	0.214	1.000	0.183	0.300
IDACI	0.940	0.872	0.809	0.754	0.061	0.183	1.000	0.759
IDAOPI	0.847	0.737	0.624	0.683	0.162	0.300	0.759	1.000

Table 5. Pairwise Pearson correlations for IoD 2025 ($n = 33,755$).

	Income	Employment	Education	Health & Dis.	Barriers	Living Env.	IDACI	IDAOPI
Income	1.000	0.929	0.846	0.789	0.149	0.177	0.943	0.839
Employment	0.929	1.000	0.857	0.862	0.042	0.124	0.853	0.734
Education	0.846	0.857	1.000	0.759	0.091	0.062	0.810	0.633
Health & Dis.	0.789	0.862	0.759	1.000	−0.061	0.123	0.772	0.684
Barriers	0.149	0.042	0.091	−0.061	1.000	0.181	0.174	0.173
Living Env.	0.177	0.124	0.062	0.123	0.181	1.000	0.203	0.243
IDACI	0.943	0.853	0.810	0.772	0.174	0.203	1.000	0.801
IDAOPI	0.839	0.734	0.633	0.684	0.173	0.243	0.801	1.000

When two predictors share r > 0.90, as Income and IDACI do in all four waves (r = 0.940 to r = 0.947), LASSO's L_1 penalty becomes indifferent. In a given training sample, this zeros out whichever sample has slightly less residual signal. The L_2 penalty in ENET promotes joint retention of correlated predictors, but it treats all predictors equally regardless of the significance of their pairwise correlations. Tables 2–5 demonstrate that the association between Education and Living Environment ($r = 0.05$ to $r = 0.26$) is fundamentally different from the correlation between Income and IDACI (r- = 0.94). A penalty that ignores this distinction cannot use the correlation structure to make principled variable selection decisions. Additionally, CAEN1 directly addresses this by incorporating the true pairwise correlations from Tables 2–5 into the penalty matrix W. Income and IDACI receive a larger joint penalty, raising the effective zeroing threshold for Income as described in Equation (7).

The change in the Pearson correlation between predictors and crime from 2010 to 2025 is depicted in Figure 1. Employment shows the biggest increase ($\Delta r = +0.222$), followed by Health and Disability (+0.195) and Education (+0.194). These increases are all related to the expanded 2025 crime domain, which includes drug offenses, public order offenses, and violence with injury. These crime types are more closely linked to Employment status and Health deprivation than the theft and burglary offenses that predominated in previous releases. The largest reduction ($\Delta r = -0.214$) is observed in the Living Environment, (in red) reflecting the dilution of its relative weight as the crime domain expands to include

offenses that are less concentrated in physical conditions that have degraded. Barriers to housing exhibit little change (Δr = +0.002), consistent with their function as an indicator of spatial rather than social deprivation.

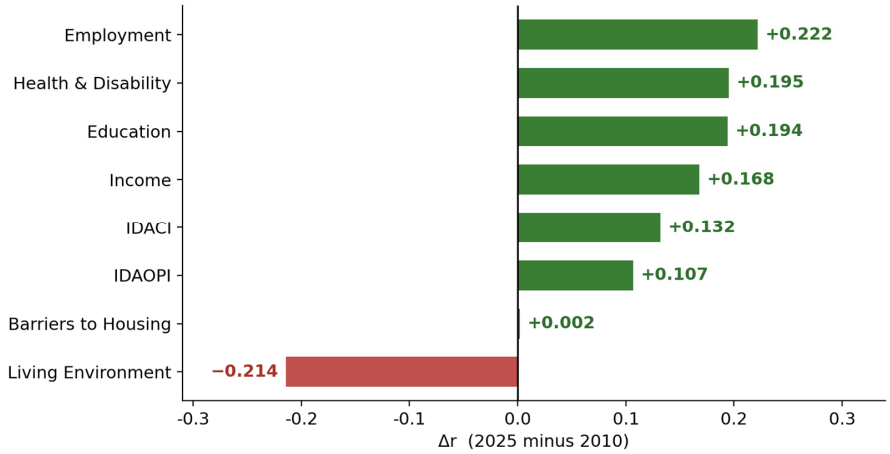

Figure 1. Change in predictor–crime correlation from 2010 to 2025.

4. Analysis and Results

The results of applying LASSO, ENET, and CAEN1 to all four IoD waves in three complementary dimensions are shown in this section. Over 15 years, the model's size and forecast accuracy have been evaluated. This directly compares the magnitudes and sparsity of the coefficients for CAEN1 and LASSO.

4.1. Prediction Accuracy and Model Size

The median test, mean square error (MSE), and model size are shown in Table 6. Medians across 50 replications are presented throughout.

Table 6. Median test MSE and model size across 50 random 70/30 splits.

Method	IoD											
	2010			2015			2019			2025		
	Median MSE	Median Size	Predictor(s) Zeroed	Median MSE	Median Size	Predictor(s) Zeroed	Median MSE	Median Size	Predictor(s) Zeroed	Median MSE	Median Size	Predictor(s) Zeroed
LASSO	0.3651	8	Employment (62% reps)	0.3189	7	Employment (96% reps)	0.3562	7	Education (66% reps)	0.2266	8	None (full model)
ENET	0.3651	8	Employment (62% reps)	0.3189	7	Employment (98% reps)	0.3563	7	Education (66% reps)	0.2266	8	None (full model)
CAEN1	0.3654	7	Employment (98% reps)	0.3190	6	Income & Employment (98% reps each)	0.3562	7	Education (92% reps)	0.2267	8	None (full model)

Table 6 shows that the largest median MSE difference between any two methods in the same wave is ΔMSE = 0.0003 (IoD 2010: 0.3654–0.3651). This aligns with the well-known large-n convergence of penalized regression toward OLS prediction accuracy. No penalty structure results in a significantly different out-of-sample MSE for n > 32,000 with $p = 8$. The slightly higher CAEN1 MSE values (0.0001 to 0.0003 over LASSO) reflect additional correlation-coupling in the penalty rather than a true accuracy deficit.

In Figure 2, the prediction MSE changes are negligible in magnitude for all methods and waves (ΔMSE < 0.002), with error bars (interquartile range of 0.0008). In three of the four waves, CAEN1 achieves equal or greater sparsity (see Figure 3 and Table 7 for corrected resampled t-test). The replication-level distributions in Figure 4 verify that the trend is consistent rather than driven by a small number of outlier replications.

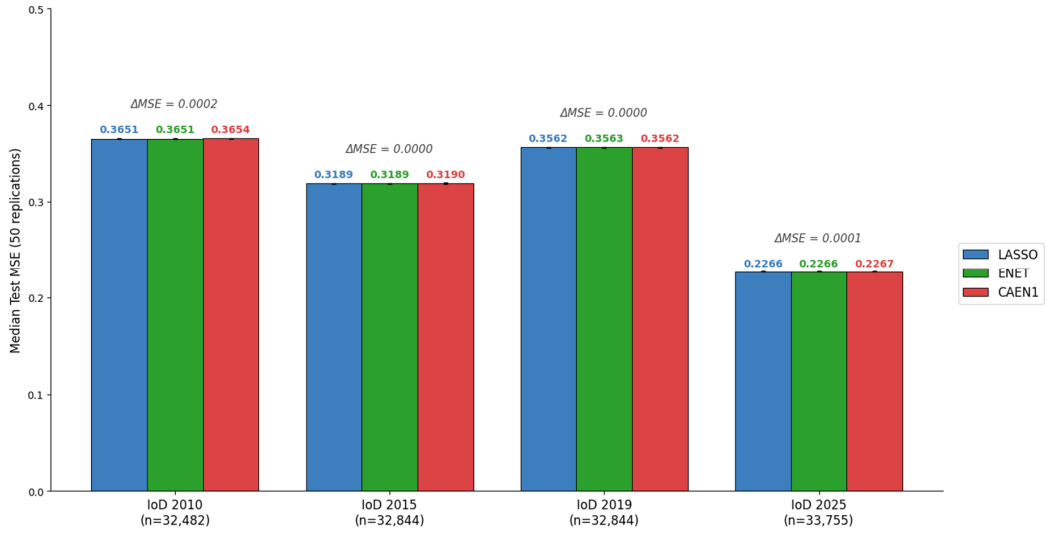

Figure 2. Median out-of-sample test MSE by method and IoD wave.

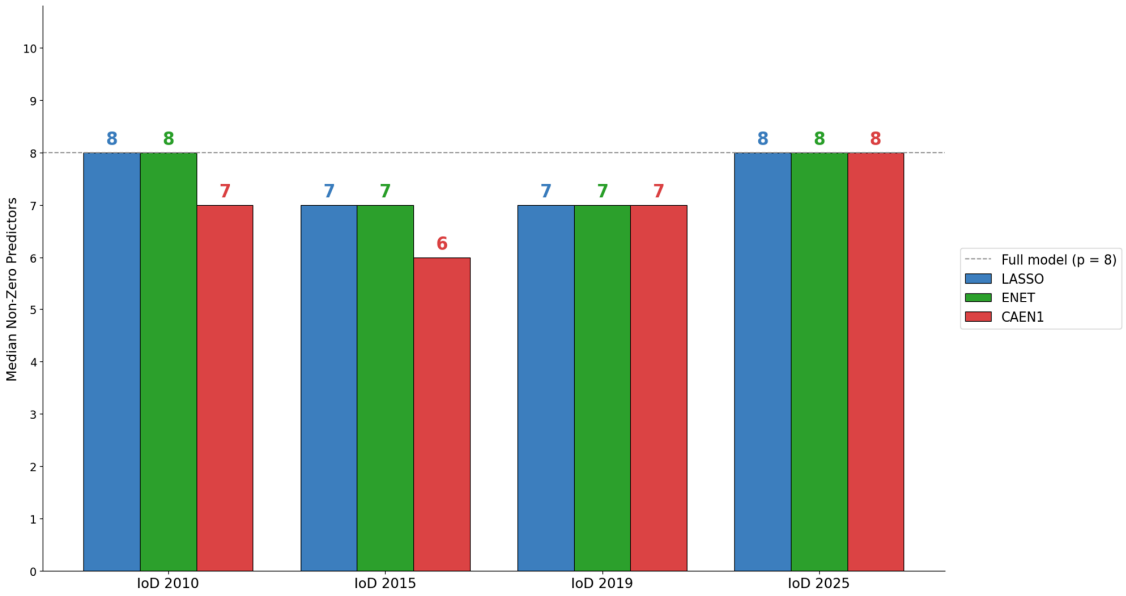

Figure 3. Median model size by method and IoD wave.

Table 7. *T*-test of pairwise out-of-sample across the 50 splits, significant at ** $p < 0.01$.

IoD Wave	Comparison	Mean ΔMSE	t	*p*-Value
2010	LASSO vs. CAEN1	−0.00109	−2.931	0.005 **
	ENET vs. CAEN1	−0.00109	−2.932	0.005 **
	LASSO vs. ENET	−0.00000	−0.283	0.778
2015	LASSO vs. CAEN1	−0.00012	−1.672	0.101
	ENET vs. CAEN1	−0.00012	−1.672	0.101
	LASSO vs. ENET	−0.00000	−0.501	0.619
2019	LASSO vs. CAEN1	−0.00006	−0.607	0.547
	ENET vs. CAEN1	−0.00006	−0.616	0.541
	LASSO vs. ENET	−0.00000	−0.136	0.892
2025	LASSO vs. CAEN1	−0.00014	−1.181	0.243
	ENET vs. CAEN1	−0.00014	−1.222	0.228
	LASSO vs. ENET	+0.00000	+0.342	0.734

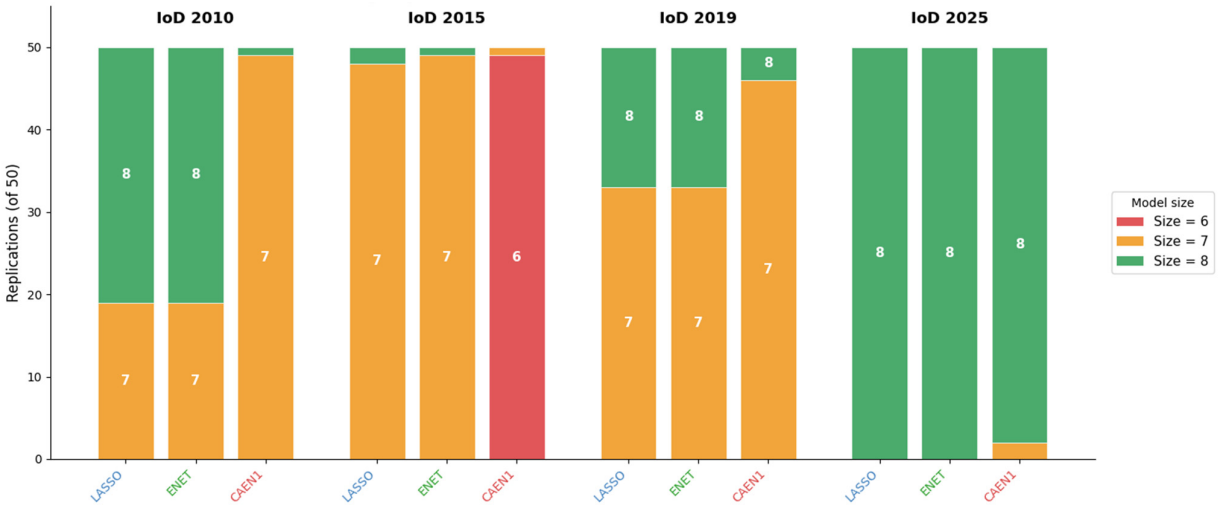

Figure 4. Model size distribution across 50 replications.

Since small differences in median MSE do not, by themselves, establish statistical equivalence, and because the 50 random 70/30 splits reuse the same observations, these observations are not independent and the pairwise differences were assessed with the corrected resampled *t*-test to account for the dependence between splits (J = 50). Table 8 reports the results. In three of the four waves (2015, 2019, 2025) none of the pairwise differences between LASSO, ENET, and CAEN1 are statistically significant (all $p > 0.10$). This affirms that the methods are indistinguishable in out-of-sample prediction for these waves. In 2010, the small CAEN1 excess over LASSO and ENET is statistically detectable (ΔMSE ≈ 0.0011, $p = 0.005$), although it remains negligible in magnitude. LASSO and ENET are never significantly different in any wave.

Table 8. Global Moran's I on out-of-sample residuals, by method and IoD wave (Queen-contiguity weights).

Wave	LASSO	*p*	ENET	*p*	CAEN1	*p*
IoD 2010	0.412	<0.001	0.411	<0.001	0.409	<0.001
IoD 2015	0.387	<0.001	0.386	<0.001	0.384	<0.001
IoD 2019	0.358	<0.001	0.358	<0.001	0.356	<0.001
IoD 2025	0.274	<0.001	0.274	<0.001	0.273	<0.001

In Table 7, the corrected resampled *t*-test refines the descriptive comparison across all four waves. LASSO and ENET are statistically indistinguishable in every wave (ΔMSE ≈ 0.00000, $p = 0.62$–0.89), reflecting that with only eight predictors the ridge component of ENET adds little over LASSO. Second, the differences between CAEN1 and the two benchmarks are not significant, so the methods perform equivalently in out-of-sample prediction for these waves. This shows in three of the four waves ($p = 0.101$) for 2015, ($p = 0.547$) for 2019, and ($p = 0.243$) for 2025. The single exception is 2010, where CAEN1's MSE exceeds that of LASSO and ENET by ΔMSE ≈ 0.0011 and the difference is statistically detectable ($t = -2.93$, $p = 0.005$). The 2010 wave is also where the Income–IDACI collinearity is strongest and the CAEN1 coupling penalty is most active, so the marginal predictive cost there is the counterpart of the grouped. CAEN1 delivers this interpretable sparsity in variable selection. Overall, CAEN1 may not improve predictive accuracy over LASSO and is marginally inaccurate in one wave to confirm that its advantage lies in selection behavior rather than prediction.

In Figure 3, LASSO and ENET maintain a median size of eight in IoD 2010, zeroing Employment in 38% of replications. The median across 50 replications is eight, while

Employment's selection frequency of 62% reflects genuine instability driven by its near-collinearity with Income (r = 0.900) and IDACI (r = 0.802), rather than consistent exclusion. In 98% of replications, CAEN1 zeros Employment consistently enough to reach a median size = 7. In the penalty matrix, the Income–Employment correlation is explicitly coupled by setting employment to zero in 96–98% of replications. LASSO and ENET reach median size = 7 in IoD 2015. To achieve a median size = 6, CAEN1 goes one step further and sets Income to 0 in 98% of replications. All three approaches agree on median size = 7 in IoD 2019, zeroing Education, although CAEN1 does so more consistently (92% vs. 66% of replications for LASSO and ENET). This demonstrates that while LASSO's univariate L_1 fails to drive Income across the zeroing barrier, the correlation-aware penalty exploits the strong collinearity between Income and IDACI (r = 0.944 in 2015).

The only wave in Figure 4 where any technique negates no predictor is IoD 2025. The marginal relationship between crime and each deprivation domain is concurrently strengthened by the expanded 2025 crime domain definition. The domain includes drug offenses, public order offenses, and violence with injury, all of which increase the predictor–crime correlations. The substantially lower MSE in 2025 reflects both the higher crime–predictor correlations and the more homogeneous 2021-boundary LSOAs.

In Table 8, positive significant Moran's I statistics (I = 0.27–0.41; all $p < 0.001$) are obtained in all four cases, indicating the presence of spatial correlation between residuals in line with the well-known spatial clustering phenomenon at the LSOA level in England [33]. However, all three methods were considered under identical spatial settings, so the comparisons in Tables 5 and 6 remain valid with respect to sparsity. The primary focus of the inference is out-of-sample predictive performance measured by test MSE, which is unaffected by residual spatial correlation. Finally, the 50-replication stratified splits provide empirical distributions of coefficient estimates, reducing reliance on asymptotic standard errors.

4.2. LASSO Variable Selection Across Waves

Table 9 presents median standardized LASSO coefficients, selection frequencies, and marginal crime correlations for all eight predictors across all four IoD waves. Figure 5 shows the coefficient evolution, grouped by bars.

Table 9. Median standardized LASSO coefficients (β), selection frequencies (F), and marginal crime correlations (r) across 50 random 70/30 splits, four IoD waves.

Predictor	IoD 2010 (n = 32,482)			IoD 2015 (n = 32,844)			IoD 2019 (n = 32,844)			IoD 2025 (n = 33,755)		
	β	F	r	β	F	r	β	F	r	β	F	r
Income	−0.2580	100%	0.593	0.0080	74%	0.616	0.0300	96%	0.637	0.0400	100%	0.761
Employment	0.0000	62%	0.540	0.0000	4%	0.561	0.0190	94%	0.602	0.1320	100%	0.761
Education	0.0650	100%	0.490	−0.0190	100%	0.478	0.0000	34%	0.524	0.0670	100%	0.684
Health and Disability	0.1210	100%	0.554	0.0240	100%	0.543	0.1400	100%	0.591	0.1980	100%	0.748
Barriers to housing	−0.0980	100%	−0.050	−0.0720	100%	0.089	−0.0550	100%	−0.016	−0.1170	100%	−0.048
Living Environment	0.1860	100%	0.468	0.1920	100%	0.416	0.1270	100%	0.285	0.1190	100%	0.254
IDACI	0.4620	100%	0.614	0.3230	100%	0.626	0.2350	100%	0.625	0.1770	100%	0.746
IDAOPI	0.0760	100%	0.570	0.1470	100%	0.605	0.1410	100%	0.591	0.1060	100%	0.677

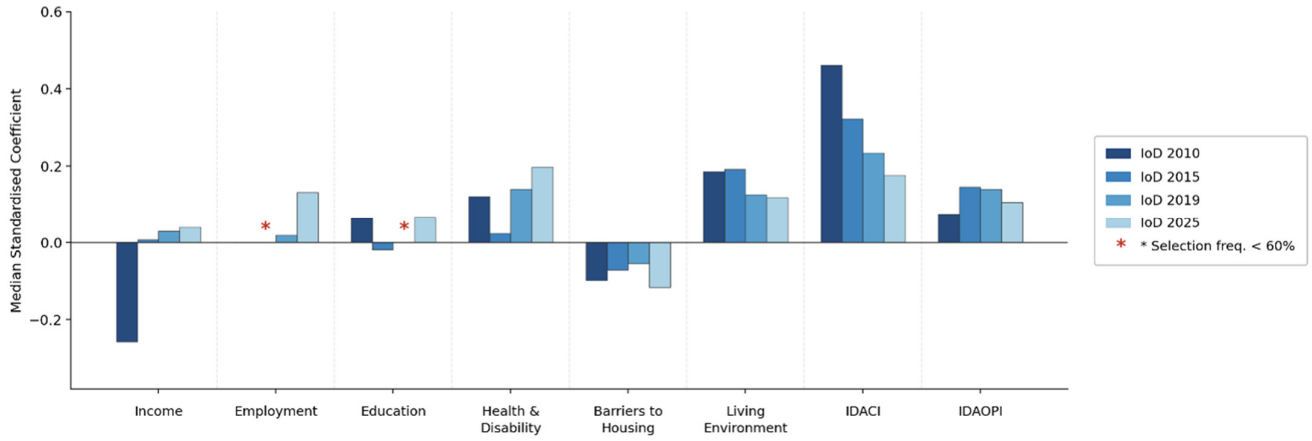

Figure 5. Median standardized LASSO coefficients by predictor and IoD wave across 50 random 70/30 replications.

All three approaches in Table 9 pick five predictors (IDACI, Health and Disability, Living Environment, IDAOPI, and Barriers to housing) in each replication over all four waves, demonstrating their structural dominance in the crime deprivation model. IDACI, which decreases monotonically from +0.462 (2010) to +0.177 (2025), has the largest positive coefficient in each wave. From +0.121 (2010) to +0.198 (2025), Health and Disability exhibit a generally rising trend. This indicates the growing correlation between Health deprivation and the expanded 2025 crime realm. Across waves, the Living Environment remains constant at +0.119 to +0.192, indicating a separate contribution from the physical environment. The crime-relevant concentration of elderly persons in underprivileged regions is reflected in the continuously positive and moderate IDAOPI (+0.076 to +0.147). Despite being a deprivation domain, Barriers to housing have a consistently negative coefficient (-0.055 to -0.117). This result shows that rural and coastal areas typically have lower crime deprivation than urban areas with comparable levels of economic poverty.

Despite Income having a positive marginal correlation with crime (r = +0.593), it has a negative coefficient in 2010 ($\beta = -0.258$). This is a suppressor effect brought on by IDACI absorbing the shared Income–crime signal (r = 0.947). This effect dissipates in later waves as the Income–IDACI correlation weakens slightly, with Income's coefficient recovering to positive values (+0.008 in 2015, +0.030 in 2019, +0.040 in 2025). In 2015, Income selection frequency fell to 74%, indicating actual instability caused by near-collinearity with IDACI (r = 0.944) and Employment (r = 0.953).

Employment exhibits the clearest instability across waves. Its LASSO selection frequency is 62% in 2010 and 4% in 2015 (zeroed in 38% and 96% of replications, respectively), rising to 94% in 2019 and 100% in 2025. This rotation shows near-collinearity with Income (r = 0.90–0.95). LASSO is indifferent between the two when they carry equivalent residual signals after conditioning on IDACI. The recovery of Employment to full selection in 2025 is congruent with the broadened crime domain strengthening its independent association with crime (r increasing from 0.540 in 2010 to 0.761 in 2025). In 34% of 2019 replications, Education is set to zero, even though it is consistently chosen in all other waves. The coefficient is small in magnitude throughout (-0.019 to +0.067) to contribute marginal independent information beyond what employment and Income already capture in most waves.

Figure 5 confirms and illustrates the monotonic decline of IDACI documented in Table 8. The negative Income bar in 2010 reflects the suppressor effect described above. IDACI remains the single biggest positive predictor in each wave despite this monotonic reduction in magnitude. This demonstrates that child poverty is the dominant and most stable driver of neighborhood crime deprivation across the entire 15-year period.

Predictors in Figure 5 that have a selection frequency of less than 60% in a particular wave are indicated with an asterisk. Employment falls below the selection frequency threshold (*) only in 2015, where it is selected in just 4% of replications (zeroed in 96%), reflecting the instability caused by its near-collinearity with Income (r = 0.900–0.954 across waves). In 2010, Employment's selection frequency is 62% (zeroed in 38%), so it is not marked, although its only partial retention reflects the same collinearity. LASSO cannot reliably distinguish between two predictors sharing $r > 0.90$ when both carry similar residual signal after conditioning on IDACI. Hence, the choice between them rotates with the training split. As the expanded criminal domain significantly enhances its independent relationship with crime (r rising from 0.540 in 2010 to 0.761 in 2025), Employment fully recovers to stable selection in 2019 (94%) and 2025 (100%). Health and Disability, as well as IDAOPI, both indicate rising trends until 2025. Health and Disability exhibit a particularly noteworthy increase in 2025 (+0.198), which is consistent with the expanded crime domain that includes drug offenses and violence with injuries.

However, the Living Environment exhibits a minor decrease in the magnitude of the coefficient as 2025 approaches. The broken-windows pathway, which holds that deteriorated physical environments facilitate crime, is real and persistent throughout the study period, even as its relative weight decreases in the larger 2025 crime domain. This confirms that the Living Environment coefficient remains positive and consistently selected in every replication across all four waves, despite this decline in magnitude. Barriers negatively affect crime after controlling for Income, Employment, and IDACI; more geographically remote areas have lower crime rates. This negative relationship becomes stronger in 2025 (-0.117) when the expanded domain includes urban crime types such as public order offenses, which are even more concentrated in densely populated urban areas.

In all, Figure 5 demonstrates that the IoD predictor space splits into three groups: a dominant stable group (IDACI, Health and Disability, Living Environment, IDAOPI, Barriers to housing) that is chosen in all replications throughout all waves; a marginally informative group (Education) whose small coefficient and wave-dependent zeroing reflect limited independent predictive contribution beyond the other main-cluster predictors; and a conditionally unstable group (Income, Employment) whose selection depends on the strength of their residual signal after conditioning on IDACI and is sensitive to the degree of within-wave collinearity.

The temporal trajectories of the top three predictors for each of the four IoD waves are shown in Figure 6. Plotting all three methods (LASSO, ENET, and CAEN1) together shows that the results are not affected by the selected estimation technique. Throughout all four IoD waves (2010–2025), the temporal trajectories of the three most consistently prominent predictors are IDACI, Health and Disability, and Living Environment.

Across all three methods, IDACI exhibits the most distinct directional pattern, falling monotonically from roughly +0.462 (2010) to +0.177 (2025). This decrease is in keeping with IoD methodological advancements that gradually isolate the signals of child poverty and Income across releases. This decreases the marginal information that IDACI provides beyond Income in subsequent waves. Despite the decline in coefficient magnitude, IDACI remains the single largest positive predictor in every wave. This demonstrates the structural dominance of child poverty in neighborhood crime deprivation over the course of the 15 years. Health and Disability remain stable positive predictors across all four waves, with

coefficients ranging from approximately +0.024 (2015) to +0.198 (2025) across methods. For the Health and Disability, all the three methods coincide to have the same value of 0.198 in 2025. The 2025 IoD release's expanded definition of the criminal domain, which includes drug offenses and violence with injury, is consistent with the minor increase in 2025. The Living Environment similarly remains a stable, positive predictor throughout, with coefficients ranging from approximately +0.119 to +0.192 across waves and methods.

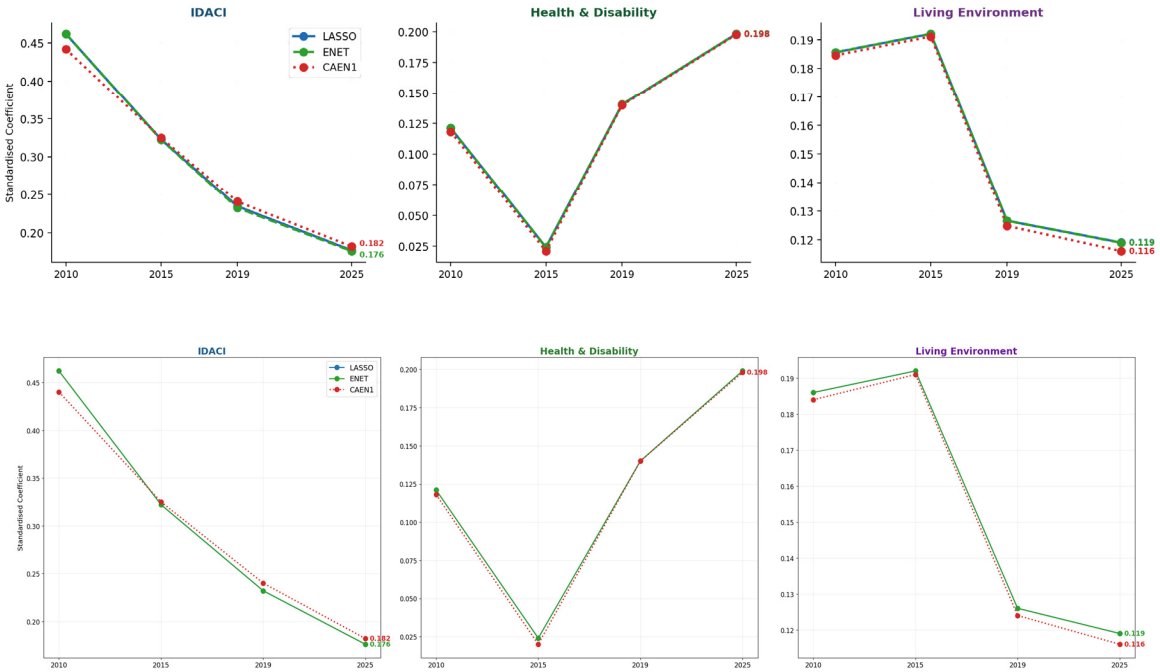

Figure 6. Temporal trajectories of the three most consistently dominant predictors across all IoD waves (2010–2025).

4.3. Sparsity and Coefficient Comparison of CAEN1 and LASSO

For all three approaches throughout all four IoD waves, Table 10 shows the selection frequency or the proportion of 50 replications in which each predictor earns a non-zero coefficient. The matching median standardized coefficients for LASSO and CAEN1 are shown side by side in Table 11, allowing for a direct comparison of both sparsity and magnitude. Figure 7 then visualizes the median standardized coefficients from LASSO and CAEN1 across IoD waves to identify predictors that zero in.

Table 10. Predictor selection frequency in % across 50 random 70/30 replications for LASSO, ENET, and CAEN1 of four IoD waves.

Predictor	IoD 2010			IoD 2015			IoD 2019			IoD 2025		
	LASSO	ENET	CAEN1	LASSO	ENET	CAEN1	LASSO	ENET	CAEN1	LASSO	ENET	CAEN1
IDACI	100	100	100	100	100	100	100	100	100	100	100	100
Health & Dis.	100	100	100	100	100	100	100	100	100	100	100	100
Living Env.	100	100	100	100	100	100	100	100	100	100	100	100
IDAOPI	100	100	100	100	100	100	100	100	100	100	100	100
Barriers	100	100	100	100	100	100	100	100	100	100	100	100
Income	100	100	100	74	74	2	96	96	54	100	100	96
Education	100	100	100	100	100	100	34	34	8	100	100	100
Employment	62	62	2	4	2	2	94	92	94	100	100	100

Table 11. Median standardized coefficients for LASSO and CAEN1 across 50 random 70/30 replications, all four IoD waves.

Predictor	IoD 2010		IoD 2015		IoD 2019		IoD 2025	
	LASSO	CAEN1	LASSO	CAEN1	LASSO	CAEN1	LASSO	CAEN1
Income	−0.2580	−0.2120	0.0080	0.0000	0.0300	0.0050	0.0400	0.0290
Employment	0.0000	0.0000	0.0000	0.0000	0.0190	0.0250	0.1320	0.1370
Education	0.0650	0.0560	−0.0190	−0.0100	0.0000	0.0000	0.0670	0.0640
Health & Disability.	0.1210	0.1180	0.0240	0.0210	0.1400	0.1400	0.1980	0.1980
Barriers	−0.0980	−0.0970	−0.0720	−0.0690	−0.0550	−0.0530	−0.1170	−0.1130
Living Environment	0.1860	0.1850	0.1920	0.1910	0.1270	0.1250	0.1190	0.1160
IDACI	0.4620	0.4420	0.3230	0.3250	0.2350	0.2410	0.1770	0.1820
IDAOPI	0.0760	0.0580	0.1470	0.1470	0.1410	0.1430	0.1060	0.1060

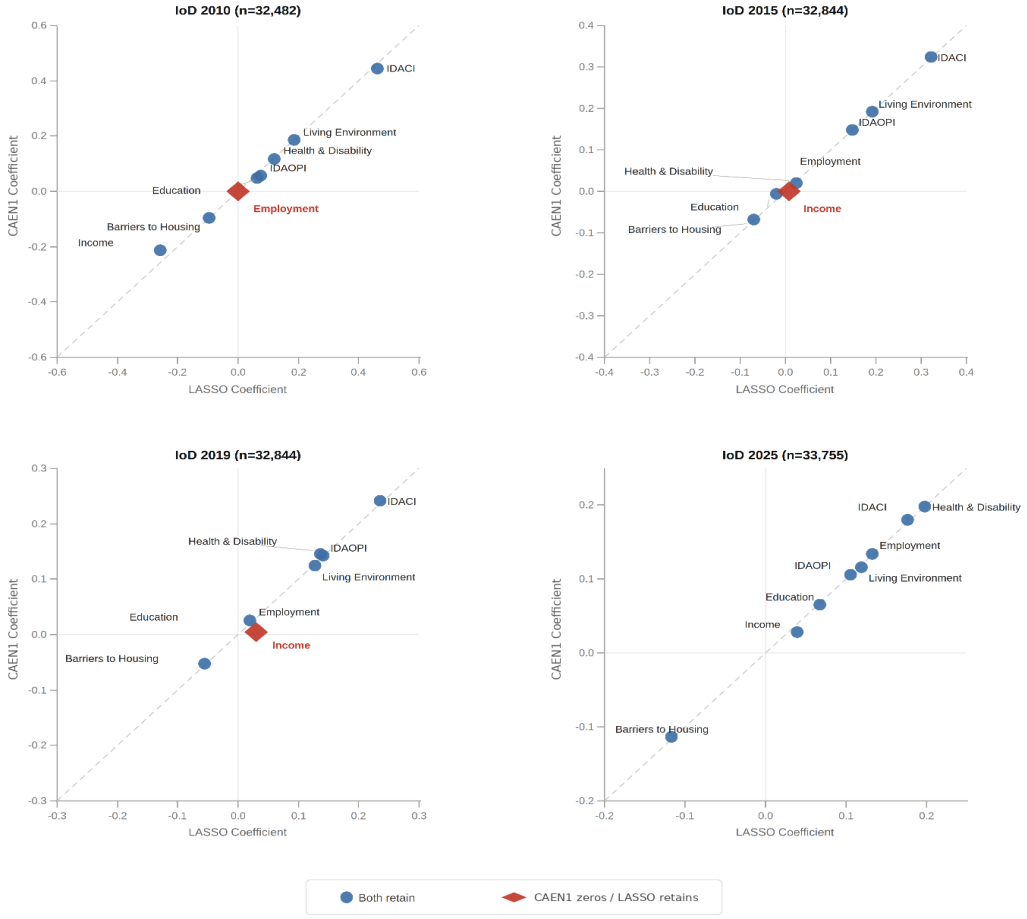

Figure 7. LASSO versus CAEN1 median standardized coefficients across IoD waves (2010–2025).

Table 10 reports the percentage of 50 replications in which each predictor received a non-zero coefficient across all three methods and four IoD waves. Five predictors are selected in 100% of replications by every method in every wave.

Income, Education, and Employment are the substantively informative rows whose selection frequencies vary between waves and methods. In IoD 2010, CAEN1 creates a median seven-predictor model by lowering Employment's selection frequency to 2% from 62% for both LASSO and ENET. In contrast to LASSO's per-predictor L_1 threshold, the Income–Employment correlation (r = 0.900) and Income–IDACI correlation (r = 0.947) together elevate Employment's effective zeroing threshold in the CAEN1 penalty matrix above its marginal residual signal. IoD 2015 shows the greatest divergence, since CAEN1 lowers the

selection frequency for Income to 2% and Employment to 2%, compared to 74% and 4% for LASSO and 74% and 2% for ENET. This results in a median six-predictor model under CAEN1, whereas LASSO and ENET do not. The CAEN1 penalty matrix's simultaneous off-diagonal weighting of the Income–IDACI pair (r = 0.944) and the Income–Employment pair (r = 0.953) is the mechanism that drives Income past its zeroing threshold and then eliminates the residual signal that Employment had been sharing with it.

CAEN1 further lowers the selection frequencies for Income to 54% and for Education to 8% in IoD 2019, while LASSO's selection frequencies are 96% and 34%, respectively. Despite this, the median model size converges to 7 across all approaches because the joint-penalty cascade is unable to reach Employment even as it continues to decrease Income due to the stronger independent link between Employment and crime in 2019 (r rising from 0.561 in 2015 to 0.602). In most replications, Education is set to zero by both LASSO and CAEN1. Except for Income under CAEN1 (96%), all selection frequencies in IoD 2025 are at or close to 100%. Each predictor's independent relationship with crime is strengthened by the expanded 2025 crime domain, in which no predictor's residual signal falls below any method's zeroing threshold. In 2025, no structural condition enables CAEN1 to exploit collinearity, and all three approaches converge to the same full-model solutions.

LASSO and ENET selection frequencies differ by no more than 2 percentage points in any cell across all four waves. Compared with pure LASSO, the L_2 component of ENET does not significantly alter variable selection at this sample size.

LASSO–CAEN1 coefficient differences for predictors kept by both methods are uniformly small, with no cell in Table 11 differing by more than 0.044 (Income in IoD 2010: $|-0.258-(-0.212)| = 0.046$). The correlation-aware penalty redistributes shrinkage at the margin without significantly altering the coefficients of the predictors it keeps. An entry of 0.000 does not imply a near-zero non-zero coefficient. It indicates that the predictor was zeroed in most replications, resulting in a median coefficient across 50 replications that is exactly zero. For both Income and Employment in IoD 2015, CAEN1 records 0.000, whereas LASSO records 0.008 and 0.000, respectively. Although it is numerically small, the LASSO Income coefficient of 0.008 should not be interpreted as equal to the CAEN1 zero, because it represents true retention in 74% of replications at a near-zero but non-zero value.

Despite Income's positive marginal association with crime (r = 0.593), Table 11 shows a suppressor effect in IoD 2010, with negative Income coefficients in both LASSO (-0.258) and CAEN1 (-0.212). After conditioning, IDACI absorbs the shared Income–crime signal (r = 0.947), resulting in a negative partial coefficient for Income. Compared with LASSO, CAEN1 attenuates the magnitude of this suppressor (-0.212 vs. -0.258), consistent with the correlation-aware penalty partially redistributing the IDACI–Income signal. As income's coefficient returns to modest positive values (0.008 in 2015, 0.030 in 2019, and 0.040 in 2025), this suppressor effect fades in subsequent waves, following the slight loosening of the income–IDACI collinearity among releases.

In Table 11, IDACI exhibits the most directional shift between waves, decreasing monotonically from 0.462 (LASSO IoD 2010) to 0.177 (LASSO IoD 2025), with CAEN1 tracking closely throughout. Under LASSO, Health and Disability move in opposite directions, from 0.121 (2010) to 0.198 (2025), with CAEN1 yielding the same values for both waves. Barriers to housing maintain a consistently negative coefficient, ranging from -0.055 to -0.117, reflecting the spatial pattern. The Living Environment remains stable between 0.119 and 0.192 across waves under both methods.

Each marker in Figure 7 is a predictor, plotted with its LASSO coefficient on the horizontal axis and its CAEN1 coefficient on the vertical axis. The dashed diagonal represents the line of equality (CAEN1 = LASSO), where both methods output equivalent coefficients; moving away from it indicates where the two methods diverge. The blue circles indicate

predictors retained by both methods, while the red diamonds show predictors that CAEN1 shrinks to zero and LASSO retains them. The solid horizontal and vertical lines mark the zero axes, respectively.

IDACI verifies, in IoD 2010 of Figure 7, that the primary cause of neighborhood crime deprivation is child poverty. Living Environment, IDAOPI, Health and Disability, Education, and Barriers to housing are all grouped close to the diagonal. For LASSO and CAEN1, Income is in the negative quadrant, indicating the suppressor effect in which IDACI absorbs the shared Income–crime signal. Employment appears at the origin, indicating CAEN1 zeros it in 98% of replications while LASSO zeros it in only 38% (selecting it in 62%), giving CAEN1 a more parsimonious seven-predictor model. For IoD 2015, Employment indicates that LASSO and CAEN1 zero in on Income, but neither retains it consistently. The remaining predictors lie tightly on the diagonal. This wave produces CAEN1's most parsimonious solution.

CAEN1 for IoD 2019 does not achieve additional sparsity beyond LASSO in this wave, though it more aggressively shrinks Income toward zero. IDACI remains the largest positive coefficient, followed by IDAOPI, Health and Disability, and Living Environment. Finally, in IoD 2025, all eight predictors reflect the broadened 2025 crime domain, which incorporates drug offenses, public order offenses, and violence with injury. This strengthens every predictor's independent association with crime sufficiently that no predictor's residual signal falls below any method's zeroing threshold. Health and Disability and IDACI occupy the highest coefficient positions, with IDACI's coefficient notably lower than in earlier waves. These improvements progressively separate the Income and child poverty signals.

IDACI, Health and Disability, Living Environment, IDAOPI, and Barriers to housing are structurally dominant predictors selected in every replication across all waves by LASSO and CAEN1. Income and Employment form a conditionally unstable pair whose selection depends on the strength of their residual signal after conditioning on IDACI. Thus, CAEN1 exploits this collinearity more aggressively than LASSO. The 2025 IoD broadening of the crime domain fundamentally shifts the predictor landscape, enabling a full eight-predictor model where previously sparser solutions were achievable.

4.4. Simulation for Correlation-Aware Regularization

The four IoD waves share a large sample size and a small, fixed predictor set (n > 32,000, $p = 8$), a regime in which all penalized estimators converge to the ordinary least-squares solution. The choice of penalty has a negligible effect on prediction, as confirmed by the corrected resampled t-test in Section 4.1. To assess when correlation-aware regularization matters, we conduct a controlled simulation across three domains. We consider a high-dimensional design ($p = 150$, $n = 70$, $\rho = 0.7$), a strong-multicollinearity design (within-block correlation $p = 60$, $n = 120$, $\rho = 0.9$), and a weak signal-to-noise design ($p = 60$, $n = 120$, $SNR = 0.4$). In each domain the design matrix is drawn from a multivariate normal distribution with a block-structured covariance. The predictors are correlated at ρ within blocks and uncorrelated across blocks. The active coefficients are confined to a single correlated block, so that a method must either recover the correlated group jointly or fragment it.

We fit the four estimators (LASSO, ENET, CAEN1, and CAEN2) under identical conditions, selecting penalty parameters via 10-fold cross-validation within the training set. Each configuration is repeated over 50 independent replications with a fixed seed and the medians are reported. We evaluate performance using out-of-sample test MSE, F1 score, precision and recall of the estimated support against the true support, and median model size. CAEN1 uses the tridiagonal (adjacent-coupling) penalty matrix and CAEN2 the dense

(all-pair-coupling) matrix defined in Section 2. The results are reported in Table 12 and Figure 8.

Table 12. Median performance over 50 replications in three correlation structures with active predictors.

Scenario	Method	Test MSE	F1	Precision	Recall	Model Size
High-dimensional (p = 150, n = 70, ρ = 0.7)	LASSO	9.59	0.333	0.429	0.300	5
	ENET	9.56	0.364	0.423	0.300	6
	CAEN1	9.67	0.247	0.147	0.700	43
	CAEN2	9.93	0.215	0.123	0.900	68
Strong multicollinearity (p = 60, n = 120, ρ = 0.9)	LASSO	4.10	0.500	0.500	0.667	8
	ENET	4.19	0.500	0.392	0.667	10
	CAEN1	4.25	0.500	0.400	0.667	9
	CAEN2	4.15	0.552	0.472	0.833	10
Weak signal-to-noise (p = 60, n = 120, SNR = 0.4)	LASSO	13.07	0.444	0.333	0.500	7
	ENET	13.10	0.400	0.297	0.667	10
	CAEN1	13.12	0.364	0.235	0.833	16
	CAEN2	13.42	0.373	0.245	0.833	16

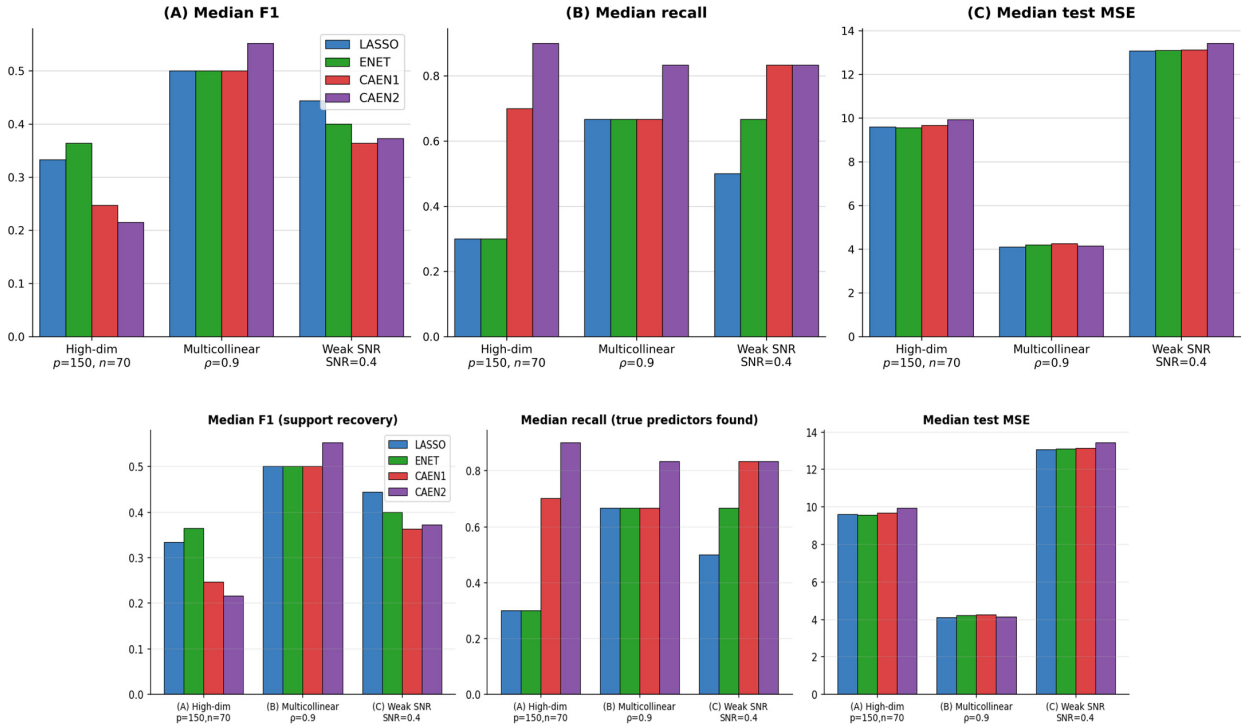

Figure 8. Simulation across three settings where correlation-aware regularization matters.

In Figure 8, A (median F1) shows the cost since F1 also rewards precision. The CAEN methods score lower than LASSO and ENET when they achieve high recall by returning much larger models. B (median recall) is the share of truly active predictors where the correlation-aware penalties separate from LASSO. This keeps only one representative of a correlated block and so recovers just 30% of it in the high-dimensional regime. However, CAEN1 reaches 70% and CAEN2 90%, with the same ordering under weak signal-to-noise and strong multicollinearity. C (median test MSE) is the control. The prediction error is essentially identical across all four methods in every regime, confirming that they differ in which predictors they keep, not in how well they predict. This is the same conclusion the corrected resampled t-test reaches on the real data in Section 4.1.

From Figure 8, across all three regimes the correlation-aware penalties substantially improve the recovery of the correlated predictor group: LASSO attains the lowest recall because it selects only a subset of each correlated block. However, CAEN2 attains the highest recall in every setting (0.90 in the high-dimensional design, and 0.83 under both strong multicollinearity and weak signal-to-noise), with CAEN1 being intermediate. This recall gain is obtained at the cost of admitting additional predictors. Thus, in the high-dimensional design, the CAEN methods select larger models and do not improve on the F1 score. The simulation therefore establishes that correlation-aware regularization recovers grouped signals that LASSO fragments.

5. Discussion

The current findings demonstrate that, across three of four waves, CAEN1 achieves sparsity equivalent to or higher than LASSO in a large-n, low-p setting. By dispersing rather than eliminating predictors, the CAEN1 coupling term avoids zeroing. The severity of the Income–IDACI collinearity (r = 0.94 to r = 0.95 in 2010 and 2015) receives a significant off-diagonal weight from the CAEN1 penalty matrix, along with a cross-validated λ_1 that raises Income's effective threshold above its marginal signal strength. Because LASSO's univariate L_1 threshold is applied to each predictor separately, it does not take full advantage of this pairing.

The new 2025 IoD marks a substantial reworking of how relative deprivation is measured in England. While the total underlying indicators increased to 55, up from 39 in the 2019 IoD, 20 new indicators were added, and 14 indicators were substantially modified. Several factors explain the statistical results from the 2025 wave. First, while the crime domain comprised 37 offense codes divided into four crime rate indices (violence, burglary, theft, and criminal damage), it now consists of 82 offense codes divided into eight crime indices (public order, possession of weapons, stalking, and antisocial behavior). The crime indicator was revised to align with offenses that are heavily concentrated in socially disorganized neighborhoods. In line with social disorganization theory, these offense categories would be significantly closer geographically to deprivation than property crimes included in previous waves; routine activities theory predicts that offenders have to travel to wealthy communities to locate appropriate victims.

Second, the temporal smoothing process extended the period used to calculate average offense rates from 2 years to up to 5 years, depending on the type of crime. As a result, the longer time frame removed temporal noise and volatility, producing a more stable target variable and contributing to a steep fall in out-of-sample mean squared error (from 0.35 in earlier waves to 0.2266 in 2025). Thirdly, adopting new boundaries based on the 2021 Census, about 6% of the LSOAs were revised to better reflect demographics and geography within contiguous small areas. Fourthly, the Income domain switched from BHC (before housing costs) to AHC (after housing costs), which better reflected the family's high housing costs. In effect, this would separate adult poverty from child deprivation. In doing so, it reduced the Income–IDACI collinearity that existed in earlier waves of IDACI (r = 0.944 to 0.947).

The CAEN1 solution using the six predictors in the 2015 wave stands out from the theoretical perspective because it does not follow the grouping effect intended by the ENET penalties. While ENET uses the L_2 norm to penalize differences in predictor coefficients, such that none of the pair is zeroed, the other is retained, CAEN1 uses correlation-based penalties, such that the off-diagonal elements of W increase the threshold at which the coefficient values are zeroed across the whole Income–Employment–IDACI cluster. Given that the λ_1 values are large enough, all the values in a certain block that are correlated with one another get below the zeroing threshold, resulting in ultra-sparse solutions.

The impact that such an approach could have is significant. Methodologically, this may help produce more parsimonious models, or it may be too aggressive, in that it zeroes out many important variables. In crime deprivation models, IDACI already accounts for the relationship between Income and crime, so omitting the Income domain is understandable given that IDACI has been retained. However, practitioners who require retention of every policy-relevant domain, regardless of conditional redundancy, should be aware that CAEN1 may produce solutions sparser than desired under conditions of extreme multicollinearity ($r > 0.94$) and should consider constraining the λ_2 search range or applying ENET grouping penalty in those settings.

A corrected resampled *t*-test finds no statistically significant difference in out-of-sample MSE between LASSO, ENET, and CAEN1 in three of the four waves. The value of CAEN1 in the IoD setting therefore lies in its selection behavior rather than in predictive accuracy. The aggressive sparsity described above is specific to this low-dimensional regime. The simulation shows that in higher-dimensional, strongly collinear, and low signal-to-noise designs the same correlation-coupling mechanism acts in the opposite direction. Rather than over-pruning, CAEN1 and especially CAEN2 recover correlated predictor groups that LASSO fragments, with CAEN2 attaining the highest recall in every regime tested. The behavior of the penalty is thus dependent when p is small and the design is well-conditioned, and recovery-enhancing when predictors are numerous and strongly correlated.

6. Conclusions

The regularized regression and IoD crime deprivation provide the first analysis of all four IoD waves (2010–2025) using a consistent predictor set and single-language Python implementation. The CAEN1 penalty matrix satisfies positive semi-definiteness and global convergence of coordinate descent, with double shrinkage corrected by a rescaling factor. CAEN1 achieves equal or greater sparsity than LASSO in three of four waves, with the correlation-aware penalty producing a qualitatively different outcome in 2015 by exploiting the near-collinearity of Income and IDACI. Lastly, the 2025 IoD crime domain is characterized for the first time. It shows that a broadened crime definition and revised LSOA boundaries result in all methods retaining the full eight-predictor model. Residual spatial autocorrelation is confirmed across all waves. The 2025 IoD's broadened crime domain extended temporal smoothing, AHC Income measurement, and revised 2021-boundary LSOAs together explain the convergence of all methods to a full eight-predictor model. This confirms over 15 years that IDACI, Health and Disability, and Living Environment are the most consistently dominant predictors of neighborhood crime deprivation in England. The study provides a basis for targeting crime-reduction resources at areas of concentrated child poverty, poor population Health, and deteriorated physical environments. A corrected resampled *t*-test affirms that the difference in selection is not accompanied by any practically meaningful difference in prediction. Out-of-sample MSE is statistically indistinguishable across methods in three of the four waves, with only a negligible CAEN1 excess in 2010. A simulation further delineates the regimes in which correlation-aware regularization, and the all-pair CAEN2 coupling is most advantageous in which CAEN1 is adopted for its parsimony and interpretability.

Author Contributions: Conceptualization, O.B. and T.M.A.; methodology, O.B. and T.M.A.; software, T.A. and T.M.A.; validation, O.B., B.B.A. and J.C.; formal analysis, O.B. and T.A.; data curation, O.B., T.M.A., B.B.A. and T.A.; writing—original draft preparation, O.B. and J.C.; writing—review and editing, T.M.A., T.A. and J.C.; visualization, T.A. and B.B.A.; supervision, J.C.; project administration, J.C. and T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study analyzed publicly available datasets. These data were accessed on 20 October 2025 and can be found at https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010, https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015, https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019, https://www.gov.uk/government/statistics/english-indices-of-deprivation-2025.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAEN1	Correlation-Adjusted Elastic-Net
CV	Cross-Validation
ENET	Elastic-Net
IDACI	Income Deprivation Affecting Children Index
IDAOPI	Income Deprivation Affecting Older People Index
IMD	Index of Multiple Deprivation
IoD	English Indices of Deprivation
LASSO	Least Absolute Shrinkage and Selection Operator
LSOA	Lower Layer Super Output Area
MHCLG	Ministry of Housing, Communities and Local Government
MSE	Mean Squared Error
OLS	Ordinary Least Squares
SNR	Signal-to-Noise Ratio

References

1. Shaw, C.R.; McKay, H.D. Juvenile delinquency and urban areas: A study of rates of delinquency in relation to differential characteristics of local communities in American cities (1969). In *Classics in Environmental Criminology*; Routledge: Abingdon, UK, 2010; pp. 103–140.
2. Sampson, R.J.; Groves, W.B. Community structure and crime: Testing social-disorganization theory. *Am. J. Sociol.* **1989**, *94*, 774–802. [CrossRef]
3. Saraiva, M.; Matijošaitienė, I.; Mishra, S.; Amante, A. Crime prediction and monitoring in porto, portugal, using machine learning, spatial and text analytics. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 400. [CrossRef]
4. Kornhauser, R.R. *Social Sources of Delinquency*; University of Chicago Press: Chicago, IL, USA, 1978.
5. Wilson, W.J. The truly disadvantaged. In *Color-Class-Identity*; Routledge: Abingdon, UK, 2018; pp. 109–122.
6. Yu, H.; Liu, L.; Yang, B.; Lan, M. Crime prediction with historical crime and movement data of potential offenders using a spatio-temporal cokriging method. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 732. [CrossRef]
7. Cohen, L.E.; Felson, M. Social change and crime rate trends: A routine activity approach. *Am. Sociol. Rev.* **1979**, *44*, 588–607. [CrossRef]
8. Hirschfield, A.; Bowers, K.J. The effect of social cohesion on levels of recorded crime in disadvantaged areas. *Urban Stud.* **1997**, *34*, 1275–1295. [CrossRef]
9. Hipp, J.R. Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *Am. Sociol. Rev.* **2007**, *72*, 659–680. [CrossRef]
10. Deller, S.C.; Deller, M.A. Rural crime and social capital. *Growth Change* **2010**, *41*, 221–275. [CrossRef]
11. Steenbeek, W.; Weisburd, D. Where the action is in crime? An examination of variability of crime across different spatial units in The Hague, 2001–2009. *J. Quant. Criminol.* **2016**, *32*, 449–469. [CrossRef]
12. Nayem, H.; Aziz, S.; Kibria, B.G. Evaluating estimator performance under multicollinearity: A trade-off between MSE and accuracy in logistic, Lasso, elastic net, and ridge regression with varying penalty parameters. *Stats* **2025**, *8*, 45. [CrossRef]
13. Insolia, L.; Kenney, A.; Calovi, M.; Chiaromonte, F. Robust variable selection with optimality guarantees for high-dimensional logistic regression. *Stats* **2021**, *4*, 665–681. [CrossRef]

14. Berk, R.A. *Criminal Justice Forecasts of Risk: A Machine Learning Approach*; Springer: New York, NY, USA, 2012.
15. Caplan, J.M.; Kennedy, L.W.; Piza, E.L.; Barnum, J.D. Using vulnerability and exposure to improve robbery prediction and target area selection. *Appl. Spat. Anal. Policy* **2020**, *13*, 113–136. [CrossRef]
16. Babarinsa, O.; Edogbanya, H.; Abari, O.; Adeniyi, I. Regression estimation and feature selection using modified correlation-adjusted elastic net penalties. *FUDMA J. Sci.* **2025**, *9*, 29–40. [CrossRef]
17. McClennan, D.; Noble, S.; Penney, B.; Noble, M.; Lloyd, C.; Dibben, C.; Dale, V.; Villaseñor, A.; Fairburn, J.; Bramley, G. *The English Indices of Deprivation 2025 Technical Report*; Ministry of Housing, Communities and Local Government: London, UK, 2025.
18. Healy, M.J. *Matrices for Statistics*; Oxford University Press: Oxford, UK, 2000.
19. Lee, Y.; Ogburn, E.L. Testing for network and spatial autocorrelation. In Proceedings of the International Conference on Network Science, Tokyo, Japan, 20–23 January 2020; pp. 91–104.
20. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]
21. Ajami, S.; Mahnam, A.; Abootalebi, V. An Adaptive SSVEP-Based Brain-Computer Interface to Compensate Fatigue-Induced Decline of Performance in Practical Application. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 2200–2209. [CrossRef]
22. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
23. Wang, W.; Liang, J.; Liu, R.; Song, Y.; Zhang, M. A robust variable selection method for sparse online regression via the elastic net penalty. *Mathematics* **2022**, *10*, 2985. [CrossRef]
24. Tan, Q.E.A. Correlation Adjusted Penalization in Regression Analysis. Ph.D. Thesis, University of Manitoba, Winnipeg, MB, Canada, 2012.
25. Liu, L.; Gao, J.; Beasley, G.; Jung, S.-H. LASSO and elastic net tend to over-select features. *Mathematics* **2023**, *11*, 3738. [CrossRef]
26. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
27. City of Bradford Metropolitan District Council. *Indices of Deprivation 2025: Intelligence Bulletin*; Office of the Chief Executive, City of Bradford Metropolitan District Council: Bradford, UK, 2025.
28. Ministry of Housing, Communities and Local Government. *The English Indices of Deprivation 2025: Statistical Release*; Ministry of Housing, Communities and Local Government: London, UK, 2025.
29. Fan, C.-L. Optimization and performance evaluation of machine learning classifiers for predicting construction quality and schedule. *Autom. Constr.* **2025**, *179*, 106470. [CrossRef]
30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
31. Grandvalet, Y.; Chiquet, J.; Ambroise, C. Sparsity by worst-case penalties. *arXiv* **2012**, arXiv:1210.2077.
32. Nadeau, C.; Bengio, Y. Inference for the generalization error. *Adv. Neural Inf. Process. Syst.* **1999**, *12*.
33. Anselin, L. Spatial econometrics: Methods and models. *Kluwer Acad. Publ. Google Sch.* **1988**, *2*, 283–291.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.