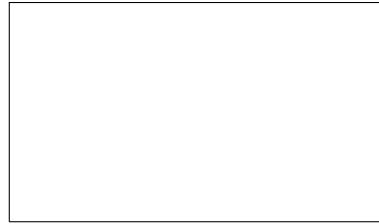


Graphical Abstract

Sparse–Smooth Spatially Varying Coefficient Quantile Regression

HOU Jian, MENG Tan, TIAN Maozai



Highlights

Sparse–Smooth Spatially Varying Coefficient Quantile Regression

HOU Jian, MENG Tan, TIAN Maozai

- Parsimonious global-local decomposition distinguishes constant effects from spatial deviations via adaptive group selection.
- Normalized Graph-Laplacian smoothing adapts to irregular sampling densities while ensuring geometric fidelity.
- Convex formulation solved by efficient ADMM and Moreau-smoothed proximal-gradient algorithms.
- Establishes oracle properties: selection consistency for local deviations and root- n normality for global coefficients.
- Outperforms kernel-based methods in recovering heterogeneous patterns under heavy-tailed distributions.

Sparse–Smooth Spatially Varying Coefficient Quantile Regression

HOU Jian

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, Beijing, China

MENG Tan

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, Beijing, China

TIAN Maozai

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, Beijing, China

Abstract

Modeling spatially varying relationships in the presence of heavy-tailed responses and irregular sampling designs presents a significant challenge in geospatial analysis. In this paper, we propose a Sparse–Smooth Spatially Varying Coefficient Quantile Regression (SS–SVCQR) framework that simultaneously achieves robust estimation, structural interpretability, and computational scalability. Our approach decomposes each coefficient function into a global baseline and a spatial deviation component. By applying a group L_2 penalty, the model automatically selects between globally constant and spatially varying specifications, offering a parsimonious alternative to standard Geographically Weighted Regression (GWR). Conditional on local variation, spatial smoothness is enforced via a normalized graph Laplacian regularization, which adapts to varying sampling densities and admits a Bayesian interpretation as an intrinsic Gaussian Markov Random Field (GMRF) prior. We develop efficient convex optimization algorithms based on ADMM and smoothed proximal-gradient methods to handle the non-differentiable check loss. Theoretically, we establish the root- n consistency and asymptotic normality of the global estimators, as well as the selection consistency of the global-vs-local structure under mild conditions. Extensive simulations and

a real-world housing application show that SS–SVCQR provides more accurate quantile predictions and more interpretable global–local structures than existing kernel- and spline-based alternatives, especially in the presence of heavy-tailed noise.

Keywords: Quantile Regression, Spatially Varying Coefficients, Graph Laplacian, Variable Selection, Group Lasso, ADMM.

1. Introduction

Spatial nonstationarity—the phenomenon where relationships between a response variable and its predictors vary across geographical locations—poses a fundamental challenge in regression analysis. Ignoring this heterogeneity by fitting a global constant-coefficient model can lead to biased estimates and misleading inference. To address this, the literature on Spatially Varying Coefficient (SVC) models has flourished. The most prominent approach is Geographically Weighted Regression (GWR) [1, 2], which fits local weighted regressions at each target location. While effective, standard GWR often suffers from multicollinearity and bandwidth selection sensitivity. Recent extensions have sought to mitigate these issues through multiscale formulations [3, 4] and mixed GWR (MGWR) that allows some coefficients to remain fixed [5, 6]. Alternative SVC frameworks include Bayesian hierarchical models based on Gaussian Processes (GPs) [7, 8] and frequentist approaches using penalized splines [9, 10]. However, GP-based methods often face cubic computational complexity, limiting their applicability to large-scale geospatial datasets.

Parallel to mean regression, Quantile Regression (QR) [11] has become an indispensable tool for characterizing the entire conditional distribution. In spatial contexts, modeling conditional quantiles is particularly valuable because spatial heterogeneity often manifests asymmetrically—affecting the tails of the distribution differently than the center [12]. Although Spatially Varying Coefficient Quantile Regression (SVC-QR) offers a powerful lens for such analysis, existing methods face dual challenges: interpretability and scalability. Many nonparametric SVC-QR estimators [13, 14] yield coefficients that vary everywhere, failing to distinguish between truly local effects and those that are effectively global. This lack of parsimony complicates interpretation. While variable selection techniques like the Lasso have been adapted for spatial models [15, 16], scalable convex formulations that si-

multaneously deliver interpretable global-vs-local selection and graph-based smoothness remain relatively limited. Recent developments in "Network Lasso" and graph-trend filtering [17, 18] suggest that graph-based penalties offer a computationally efficient alternative to kernel or spline methods, yet their integration with quantile loss in a spatial setting, together with an explicit mechanism for global-vs-local selection, remains underexplored.

In this paper, we propose a penalized estimator that addresses these gaps by integrating graph Laplacian regularization with group-sparse penalization. Our approach, termed SS-SVCQR (Sparse-Smooth Spatially Varying Coefficient Quantile Regression), is designed with two key desiderata. First, we employ a decomposition strategy that separates each coefficient into a global baseline and a spatial deviation. A group L_2 penalty [19] is applied to the deviation vector, shrinking it to zero when the covariate effect is globally constant. This provides an automated, interpretable decision mechanism for global-vs-local specification, a feature often absent in kernel-weighted local QR. Second, conditioned on being local, we regularize the deviations using a normalized graph Laplacian constructed from the sampling locations [20, 21]. This encourages spatial continuity and allows the model to adapt to irregular sampling designs without the heavy computational burden of dense covariance matrices used in GPs. The resulting formulation is convex and compatible with large-scale solvers, including Alternating Direction Method of Multipliers (ADMM) schemes tailored to the non-differentiable check loss.

On the theoretical front, we establish that our estimator enjoys desirable asymptotic properties. We show that the global coefficients are root- n consistent and asymptotically normal, while the deviation fields converge in mean square at a rate that balances Laplacian bias and stochastic variability. Furthermore, we prove that the selection of globally constant effects is consistent under a groupwise beta-min condition. Our theoretical framework accommodates random sampling locations filling a compact set and sparse graphs with uniformly bounded average degrees.

The remainder of this paper is organized as follows. Section 2 introduces the proposed model decomposition and the graph Laplacian regularization framework. Section 3 details the optimization algorithms, including an efficient ADMM implementation and tuning parameter selection via spatially blocked cross-validation. Section 4 provides the asymptotic analysis. In Section 5, we evaluate the finite-sample performance of our method against competing GWR and spline-based approaches through extensive simulations. Section 6 applies our method to a real-world analysis of housing prices in the

Lucas County, demonstrating the model’s ability to uncover heterogeneous pricing mechanisms and distributional tail dynamics. Section 7 concludes with a discussion.

2. Model

We develop a sparse–smooth spatially varying coefficient quantile regression (SS–SVCQR) at a fixed quantile $\tau \in (0, 1)$. This framework fuses the robustness of quantile regression with the flexibility of graph-based regularization to capture heterogeneous spatial dependencies.

Scalars are denoted by italic lower case, vectors by bold lower case, and matrices by bold upper case. For a vector \mathbf{v} , $\|\mathbf{v}\|_2$ denotes the Euclidean norm, and $\mathbf{1}$ is an all-ones vector of conformable length.

2.1. Data and Decomposition Structure

We observe a spatial dataset $\{(Y_i, \mathbf{Z}_i, \mathbf{X}_i, u_i) : i = 1, \dots, n\}$, where $Y_i \in \mathbb{R}$ is the response, $\mathbf{Z}_i \in \mathbb{R}^q$ represents covariates with purely global effects, $\mathbf{X}_i \in \mathbb{R}^p$ contains predictors with potentially spatially varying effects, and $u_i \in \mathcal{U} \subset \mathbb{R}^2$ denotes the spatial location. Let $Q_\tau(Y_i | \mathbf{Z}_i, \mathbf{X}_i, u_i)$ be the conditional τ -quantile. We assume the data generating process follows a varying-coefficient structure:

$$Q_\tau\{Y_i | \mathbf{Z}_i, \mathbf{X}_i, u_i\} = \mathbf{Z}_i^\top \boldsymbol{\alpha} + \sum_{j=1}^p X_{ij} \beta_j(u_i), \quad i = 1, \dots, n, \quad (1)$$

where $\boldsymbol{\alpha} \in \mathbb{R}^q$ collects the fixed global coefficients, and $\beta_j(\cdot)$ are unknown functions characterizing the spatial heterogeneity. To simultaneously perform variable selection between global and local specifications, we decompose each varying coefficient into a global baseline level and a location-specific spatial deviation:

$$\beta_j(u) = \beta_{G,j} + \delta_j(u), \quad j = 1, \dots, p. \quad (2)$$

Here, $\boldsymbol{\beta}_G = (\beta_{G,1}, \dots, \beta_{G,p})^\top \in \mathbb{R}^p$ captures the domain-wide average effect, while $\delta_j : \mathcal{U} \rightarrow \mathbb{R}$ captures local departures. This additive form allows the model to adaptively shrink $\delta_j(\cdot)$ to zero when the effect of predictor j is homogeneous, reducing the model to a standard global quantile regression.

The model is anchored at level τ by the quantile exogeneity condition:

$$\Pr \left\{ Y_i \leq \mathbf{Z}_i^\top \boldsymbol{\alpha}_0 + \sum_{j=1}^p X_{ij} \beta_{0,j}(u_i) \mid \mathbf{Z}_i, \mathbf{X}_i, u_i \right\} = \tau, \text{ almost surely for each } i. \quad (3)$$

Here $\boldsymbol{\beta}_0 = \{\beta_{0,1}(\cdot), \dots, \beta_{0,p}(\cdot)\}$ is represented by $(\boldsymbol{\beta}_{G0}, \{\delta_{0,j}(\cdot)\}_{j=1}^p)$ with $\beta_{0,j}(u) = \beta_{G0,j} + \delta_{0,j}(u)$.

2.2. Graph Construction and Spatial Regularization

Since the locations $\{u_i\}$ may be irregularly distributed, we encode the spatial domain's geometry via a proximity graph. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with vertices $\mathcal{V} = \{1, \dots, n\}$ representing the observations. We construct a sparse, symmetric adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ using a mutual k -nearest-neighbor (k -NN) rule combined with a Gaussian kernel. Specifically, for two locations u_i and u_ℓ , the edge weight is defined as:

$$A_{i\ell} = \begin{cases} \exp\left(-\frac{\|u_i - u_\ell\|_2^2}{\sigma^2}\right), & \text{if } u_\ell \in \mathcal{N}_k(u_i) \text{ or } u_i \in \mathcal{N}_k(u_\ell), \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where $\mathcal{N}_k(u)$ denotes the set of k nearest neighbors of u , and σ is a bandwidth parameter. This construction ensures the graph is connected while adapting to local sampling densities.

To enforce spatial smoothness on $\boldsymbol{\delta}_j$, we utilize the spectral properties of the graph. Let $\mathbf{D} = \text{diag}(\mathbf{A}\mathbf{1})$ be the degree matrix. We adopt the symmetric normalized Laplacian:

$$\mathbf{L}_{\text{sym}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}. \quad (5)$$

We favor \mathbf{L}_{sym} over the unnormalized Laplacian $(\mathbf{D} - \mathbf{A})$ because its spectrum is bounded in $[0, 2]$ and, crucially, it normalizes for node degree. In spatial statistics, sampling locations often exhibit clustering; the normalized Laplacian ensures that the regularization penalty is not artificially inflated in densely sampled regions [22].

The spatial roughness of the deviation vector $\boldsymbol{\delta}_j = (\delta_j(u_1), \dots, \delta_j(u_n))^\top$ is quantified by the quadratic form:

$$\boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j = \frac{1}{2} \sum_{(i,\ell) \in \mathcal{E}} A_{i\ell} \left\{ \frac{\delta_j(u_i)}{\sqrt{d_i}} - \frac{\delta_j(u_\ell)}{\sqrt{d_\ell}} \right\}^2, \quad d_i = (\mathbf{D})_{ii}. \quad (6)$$

This penalty term admits a probabilistic interpretation. It corresponds to placing an intrinsic Gaussian Markov Random Field (GMRF) prior on the spatial deviations, $p(\boldsymbol{\delta}_j) \propto \exp(-\frac{\lambda_2}{2} \boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j)$. This prior encodes the belief that, conditional on its neighbors, the deviation at a node is normally distributed around the degree-weighted average of its neighbors, thereby enforcing local coherence.

2.3. Identifiability and Matrix Formulation

The decomposition in (2) requires a constraint to prevent collinearity between the global intercept $\beta_{G,j}$ and the spatial field $\delta_j(\cdot)$. We impose that local deviations are degree-weight centered componentwise. For a connected component $\mathcal{C} \subset \{1, \dots, n\}$ with an indicator vector $\mathbf{1}_{\mathcal{C}}$, the identifiability constraint is:

$$\mathbf{1}_{\mathcal{C}}^\top \mathbf{D} \boldsymbol{\delta}_j = 0, \quad (7)$$

for all $j = 1, \dots, p$. Since the null space of \mathbf{L}_{sym} on a component is spanned by $\mathbf{D}^{1/2} \mathbf{1}_{\mathcal{C}}$, ensuring $\mathbf{D}^{1/2} \boldsymbol{\delta}_j$ is orthogonal to this null space is equivalent to (7). Under this constraint, $\beta_{G,j}$ interprets as the *degree-weighted spatial average* of the total coefficient $\beta_j(\cdot)$, while $\delta_j(\cdot)$ captures pure spatial variation.

For estimation, we vectorize the system. Stack the responses $\mathbf{y} = (Y_1, \dots, Y_n)^\top$ and design matrices $\mathbf{Z} = [\mathbf{Z}_1^\top; \dots; \mathbf{Z}_n^\top] \in \mathbb{R}^{n \times q}$ and $\mathbf{X} = [\mathbf{X}_1^\top; \dots; \mathbf{X}_n^\top] \in \mathbb{R}^{n \times p}$. To handle the varying coefficients, we define diagonal operators $\mathbf{X}_{\odot j} = \text{diag}(X_{1j}, \dots, X_{nj}) \in \mathbb{R}^{n \times n}$ for $j = 1, \dots, p$. Concatenating the deviations as $\Delta = [\boldsymbol{\delta}_1^\top, \dots, \boldsymbol{\delta}_p^\top]^\top \in \mathbb{R}^{np}$, we construct the block operator:

$$\mathbf{X}_\delta = [\mathbf{X}_{\odot 1} \ \mathbf{X}_{\odot 2} \ \cdots \ \mathbf{X}_{\odot p}] \in \mathbb{R}^{n \times np}. \quad (8)$$

so that $X_\delta \Delta$ stacks all local deviation terms in a linear operator form. The global vector of conditional quantiles is then succinctly expressed as:

$$\mathbf{q}_\tau = \mathbf{Z} \boldsymbol{\alpha} + \mathbf{X} \boldsymbol{\beta}_G + \mathbf{X}_\delta \Delta. \quad (9)$$

This linear formulation allows us to leverage efficient convex optimization algorithms to estimate the parameters $\theta = (\boldsymbol{\alpha}, \boldsymbol{\beta}_G, \Delta)$ by minimizing the check loss subject to the penalties defined above.

3. Estimation and Algorithms

We estimate $(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L)$ with $\boldsymbol{\theta}_P = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}_G^\top)^\top \in \mathbb{R}^{q+p}$ and $\boldsymbol{\Theta}_L = \{\boldsymbol{\delta}_j \in \mathbb{R}^n : j = 1, \dots, p\}$, each $\boldsymbol{\delta}_j$ degree-weight centered on every component.

The empirical objective at τ is

$$\begin{aligned} \min_{\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L} \sum_{i=1}^n \rho_\tau \left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\alpha} - \mathbf{X}_i^\top \boldsymbol{\beta}_G - \sum_{j=1}^p X_{ij} \delta_j(u_i) \right) + \lambda_1 \sum_{j=1}^p w_j \|\boldsymbol{\delta}_j\|_2 \\ + \lambda_2 \sum_{j=1}^p \boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j, \end{aligned} \quad (10)$$

subject to $\mathbf{1}_C^\top \mathbf{D} \boldsymbol{\delta}_j = 0$ for every component C and all j . Weights $w_j > 0$ enable adaptive grouping, e.g. $w_j = (\|\tilde{\boldsymbol{\delta}}_j\|_2 + a)^{-\gamma}$ with $a > 0$ and $\gamma \in (0, 1]$ [23, 24]. Tuning (λ_1, λ_2) is selected by spatially blocked cross-validation.

Define $\mathbf{r}(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = \mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}_G - \mathbf{X}_\delta\boldsymbol{\Delta}$ with $\boldsymbol{\Delta} = [\boldsymbol{\delta}_1^\top, \dots, \boldsymbol{\delta}_p^\top]^\top$. The objective (10) is convex and separable across groups for the non-smooth term.

Handling the centering constraint.. After each update of $\boldsymbol{\delta}_j$, project it to the degree-weight centered subspace. Let $\{\mathcal{C}_k\}_{k=1}^K$ be the components and $\mathbf{1}_{\mathcal{C}_k}$ their indicators. The projector reads

$$\text{Proj}_D(\mathbf{v}) = \mathbf{v} - \sum_{k=1}^K \frac{\mathbf{1}_{\mathcal{C}_k}^\top \mathbf{D} \mathbf{v}}{\mathbf{1}_{\mathcal{C}_k}^\top \mathbf{D} \mathbf{1}_{\mathcal{C}_k}} \mathbf{1}_{\mathcal{C}_k}, \quad (11)$$

and we enforce $\boldsymbol{\delta}_j \leftarrow \text{Proj}_D(\boldsymbol{\delta}_j)$ every iteration.

An ADMM scheme with closed-form proximal maps

Introduce $\mathbf{s} = \mathbf{r}(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L)$ and $\mathbf{z}_j = \boldsymbol{\delta}_j$ for $j = 1, \dots, p$, and solve

$$\min_{\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L, \mathbf{s}, \{\mathbf{z}_j\}} \sum_{i=1}^n \rho_\tau(s_i) + \lambda_1 \sum_{j=1}^p w_j \|\mathbf{z}_j\|_2 + \lambda_2 \sum_{j=1}^p \boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j \quad (12)$$

subject to $\mathbf{s} = \mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}_G - \sum_{j=1}^p \mathbf{X}_{\odot j} \boldsymbol{\delta}_j$ and $\mathbf{z}_j = \boldsymbol{\delta}_j$, together with centering via (11). Using scaled duals $(\mathbf{u}, \{\mathbf{v}_j\})$ and penalties $(\rho_s, \rho_z) > 0$, an ADMM sweep is as follows[25]:

Update of $(\boldsymbol{\alpha}, \boldsymbol{\beta}_G)$. Solve

$$\begin{bmatrix} \mathbf{Z}^\top \mathbf{Z} & \mathbf{Z}^\top \mathbf{X} \\ \mathbf{X}^\top \mathbf{Z} & \mathbf{X}^\top \mathbf{X} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta}_G \end{bmatrix} = \begin{bmatrix} \mathbf{Z}^\top (\mathbf{y} - \sum_j \mathbf{X}_{\odot j} \boldsymbol{\delta}_j - \mathbf{s} + \mathbf{u}) \\ \mathbf{X}^\top (\mathbf{y} - \sum_j \mathbf{X}_{\odot j} \boldsymbol{\delta}_j - \mathbf{s} + \mathbf{u}) \end{bmatrix}.$$

The coefficient matrix depends only on $[\mathbf{Z} \ \mathbf{X}]$ and can be factorized once.

Update of \mathbf{s} . The proximal map of the check loss is an asymmetric soft-threshold. For each i ,

$$\mathbf{s} \leftarrow \text{prox}_{\rho_s^{-1}\rho_\tau}(\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}_G - \sum_j \mathbf{X}_{\odot j} \boldsymbol{\delta}_j + \mathbf{u}), \quad (13)$$

with

$$\text{prox}_{\gamma\rho_\tau}(v) = \begin{cases} v - \gamma\tau, & v > \gamma\tau, \\ 0, & -\gamma(1-\tau) \leq v \leq \gamma\tau, \\ v + \gamma(1-\tau), & v < -\gamma(1-\tau). \end{cases}$$

Update of $\boldsymbol{\delta}_j$. Holding other blocks fixed, $\boldsymbol{\delta}_j$ solves

$$\begin{aligned} (2\lambda_2 \mathbf{L}_{\text{sym}} + \rho_s \mathbf{X}_{\odot j}^\top \mathbf{X}_{\odot j} + \rho_z \mathbf{I}) \boldsymbol{\delta}_j &= \rho_s \mathbf{X}_{\odot j}^\top (\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}_G - \sum_{\ell \neq j} \mathbf{X}_{\odot \ell} \boldsymbol{\delta}_\ell - \mathbf{s} + \mathbf{u}) \\ &\quad + \rho_z (\mathbf{z}_j - \mathbf{v}_j). \end{aligned} \quad (14)$$

Solve by sparse Cholesky or conjugate gradients; then project $\boldsymbol{\delta}_j \leftarrow \text{Proj}_D(\boldsymbol{\delta}_j)$.

Update of \mathbf{z}_j . Apply group soft-thresholding,

$$\mathbf{z}_j \leftarrow \left(1 - \frac{\lambda_1 w_j}{\rho_z \|\boldsymbol{\delta}_j + \mathbf{v}_j\|_2}\right)_+ (\boldsymbol{\delta}_j + \mathbf{v}_j), \quad j = 1, \dots, p. \quad (15)$$

Dual updates and stopping. Update

$$\begin{aligned} \mathbf{u} &\leftarrow \mathbf{u} + (\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}_G - \sum_j \mathbf{X}_{\odot j} \boldsymbol{\delta}_j - \mathbf{s}), \\ \mathbf{v}_j &\leftarrow \mathbf{v}_j + (\boldsymbol{\delta}_j - \mathbf{z}_j), \end{aligned}$$

and stop when primal residuals

$$\mathbf{r}_s = \mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}_G - \sum_j \mathbf{X}_{\odot j} \boldsymbol{\delta}_j - \mathbf{s}, \quad \mathbf{r}_{z,j} = \boldsymbol{\delta}_j - \mathbf{z}_j$$

and dual residuals

$$\mathbf{d}_s = \rho_s(\mathbf{s} - \mathbf{s}^{\text{prev}}), \quad \mathbf{d}_{z,j} = \rho_z(\mathbf{z}_j - \mathbf{z}_j^{\text{prev}})$$

satisfy $\|\mathbf{r}_s\|_2 \leq \varepsilon_{\text{pri}}$, $(\sum_j \|\mathbf{r}_{z,j}\|_2^2)^{1/2} \leq \varepsilon_{\text{pri}}$, $\|\mathbf{d}_s\|_2 \leq \varepsilon_{\text{dual}}$ and $(\sum_j \|\mathbf{d}_{z,j}\|_2^2)^{1/2} \leq \varepsilon_{\text{dual}}$, where tolerances scale with \sqrt{n} and block magnitudes. Over-relaxation and adaptive penalty updates are often beneficial.

A smoothed proximal-gradient scheme with Moreau smoothing

Define the smoothed loss $M_h(r) = \min_s\{\rho_\tau(s) + \frac{1}{2h}(s - r)^2\}$. Its gradient exists and is $1/h$ -Lipschitz:

$$\nabla_r M_h(r) = \frac{1}{h}(r - \text{prox}_{h\rho_\tau}(r)).$$

Let

$$\mathcal{G}_h(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = \sum_{i=1}^n M_h\left(r_i(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L)\right) + \lambda_2 \sum_{j=1}^p \boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j, \quad \mathcal{R}(\boldsymbol{\Theta}_L) = \lambda_1 \sum_{j=1}^p w_j \|\boldsymbol{\delta}_j\|_2.$$

Gradients follow by the chain rule:

$$\begin{aligned}\nabla_{\boldsymbol{\alpha}} \mathcal{G}_h &= -\mathbf{Z}^\top \mathbf{g}, \\ \nabla_{\boldsymbol{\beta}_G} \mathcal{G}_h &= -\mathbf{X}^\top \mathbf{g}, \\ \nabla_{\boldsymbol{\delta}_j} \mathcal{G}_h &= -\mathbf{X}_{\odot j}^\top \mathbf{g} + 2\lambda_2 \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j,\end{aligned}$$

with $\mathbf{g} = \frac{1}{h}(\mathbf{r} - \text{prox}_{h\rho_\tau}(\mathbf{r}))$. Given a step size $t > 0$ (via backtracking to ensure sufficient decrease), update

$$\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} - t \nabla_{\boldsymbol{\alpha}} \mathcal{G}_h, \quad \boldsymbol{\beta}_G \leftarrow \boldsymbol{\beta}_G - t \nabla_{\boldsymbol{\beta}_G} \mathcal{G}_h,$$

$$\boldsymbol{\delta}_j \leftarrow \text{shrink}_{t\lambda_1 w_j}(\boldsymbol{\delta}_j - t \nabla_{\boldsymbol{\delta}_j} \mathcal{G}_h), \quad \text{shrink}_\kappa(\mathbf{v}) = \left(1 - \frac{\kappa}{\|\mathbf{v}\|_2}\right)_+ \mathbf{v},$$

followed by $\boldsymbol{\delta}_j \leftarrow \text{Proj}_D(\boldsymbol{\delta}_j)$. Nesterov acceleration with standard restart is used. A continuation $h \downarrow h_{\min}$ with $\sqrt{nh} \rightarrow \infty$ preserves asymptotics.

Tuning, initialization, and implementation details

Tuning (λ_1, λ_2) and graph hyperparameters (e.g. mutual k in k -NN or kernel bandwidth) uses *spatially blocked cross-validation*. Partition locations into K spatial folds; for each held-out fold, *construct the graph and Laplacian using only the training folds*, fit the model, and evaluate the held-out check loss. This prevents leakage through cross-fold edges. Adaptive weights use a pilot fit with small λ_1 and moderate λ_2 , or a purely smooth fit with $\lambda_1 = 0$; set $w_j = (\|\tilde{\boldsymbol{\delta}}_j\|_2 + a)^{-\gamma}$ with a small a for stability. Initialize $(\boldsymbol{\alpha}, \boldsymbol{\beta}_G)$ at the global QR and set $\boldsymbol{\delta}_j = \mathbf{0}$.

Both schemes exploit sparsity. \mathbf{L}_{sym} is stored in a compressed sparse format; matrix–vector products cost $O(nd)$ with average degree d . The

multipliers $\mathbf{X}_{\odot j}$ apply in $O(n)$. In ADMM, systems in (14) have the form $\rho_z \mathbf{I} + \rho_s \text{diag}(X_{:j}^2) + 2\lambda_2 \mathbf{L}_{\text{sym}}$ and admit efficient preconditioned CG with a shared preconditioner (e.g. incomplete Cholesky of $\rho_z \mathbf{I} + 2\lambda_2 \mathbf{L}_{\text{sym}}$). In the smoothed scheme, costs are dominated by proximal evaluations (linear time) and sparse graph multiplications.

Stopping rules and KKT diagnostics

For ADMM, stopping relies on primal/dual residuals as above, with tolerances

$$\begin{aligned}\varepsilon_{\text{pri}} &= \sqrt{n}\varepsilon_{\text{abs}} + \varepsilon_{\text{rel}} \max \left\{ \|\mathbf{s}\|_2, \left\| \mathbf{y} - \mathbf{Z}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}_G - \sum_j \mathbf{X}_{\odot j} \hat{\boldsymbol{\delta}}_j \right\|_2 \right\}, \\ \varepsilon_{\text{dual}} &= \sqrt{n}\varepsilon_{\text{abs}} + \varepsilon_{\text{rel}} \max \left\{ \|\rho_s(\mathbf{s} - \mathbf{s}^{\text{prev}})\|_2, \left(\sum_j \|\rho_z(\mathbf{z}_j - \mathbf{z}_j^{\text{prev}})\|_2^2 \right)^{1/2} \right\}.\end{aligned}$$

For the smoothed scheme, stopping uses relative objective decrease and a KKT residual,

$$\|\mathbf{Z}^\top \psi_\tau(\hat{\mathbf{r}})\|_2 + \|\mathbf{X}^\top \psi_\tau(\hat{\mathbf{r}})\|_2 + \sum_{j=1}^p \text{dist}\left(\mathbf{X}_{\odot j}^\top \psi_\tau(\hat{\mathbf{r}}) + 2\lambda_2 \mathbf{L}_{\text{sym}} \hat{\boldsymbol{\delta}}_j, \lambda_1 w_j \partial \|\hat{\boldsymbol{\delta}}_j\|_2\right),$$

where $\psi_\tau(\mathbf{r})$ stacks $\psi_\tau(r_i) = \tau - \mathbf{1}\{r_i < 0\}$. The distance to the group subdifferential is

$$\text{dist}(\mathbf{v}, \lambda_1 w_j \partial \|\hat{\boldsymbol{\delta}}_j\|_2) = \begin{cases} \left\| \mathbf{v} - \lambda_1 w_j \frac{\hat{\boldsymbol{\delta}}_j}{\|\hat{\boldsymbol{\delta}}_j\|_2} \right\|_2, & \hat{\boldsymbol{\delta}}_j \neq 0, \\ \max\{\|\mathbf{v}\|_2 - \lambda_1 w_j, 0\}, & \hat{\boldsymbol{\delta}}_j = 0. \end{cases}$$

Practical choices

Penalty scales can be anchored by heuristics. A convenient grid sets λ_2 around the median of the nonzero eigenvalues of \mathbf{L}_{sym} divided by n , and λ_1 around the empirical $(1 - \tau)\tau$ score scale times $\sqrt{\log p/n}$. In noisy data, a continuation schedule that starts with larger λ_2 and decreases gradually stabilizes both solvers. When repeated locations occur, connect coincident points by a high-weight edge or merge them.

4. Asymptotic Theory

We analyze the SS–SVCQR estimator in Section 2. The parametric block is $\boldsymbol{\theta}_P = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}_G^\top)^\top \in \mathbb{R}^{q+p}$, and the local deviations are $\boldsymbol{\Theta}_L = \{\boldsymbol{\delta}_j \in \mathbb{R}^n : j = 1, \dots, p\}$, each degree-weight centered on every graph component.

Estimator and penalty.. At quantile level τ ,

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = & \sum_{i=1}^n \rho_\tau \left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\alpha} - \mathbf{X}_i^\top \boldsymbol{\beta}_G - \sum_{j=1}^p X_{ij} \delta_j(u_i) \right) + \sum_{j=1}^p \lambda_{1n} w_{j,n} \|\boldsymbol{\delta}_j\|_2 \\ & + \sum_{j=1}^p \lambda_{2n} \boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j, \end{aligned}$$

subject to $\mathbf{1}_C^\top \mathbf{D} \boldsymbol{\delta}_j = 0$ for every component C and every j . Weights $w_{j,n} > 0$ can be adaptive, e.g. $w_{j,n} = (\|\tilde{\boldsymbol{\delta}}_j\|_2 + a_n)^{-\gamma}$ with $a_n \downarrow 0$, $\gamma \in (0, 1]$.

Population target and notation.. Let $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ denote the true parameters in the sense of (3)[26]. Write $\beta_{0,j}(u) = \beta_{G0,j} + \delta_{0,j}(u)$ with degree-weight centered $\delta_{0,j}(\cdot)$ and define

$$\boldsymbol{\delta}_{j0} = (\delta_{0,j}(u_1), \dots, \delta_{0,j}(u_n))^\top, \quad \mathcal{A}_0 = \{j : \boldsymbol{\delta}_{j0} = \mathbf{0}\}, \quad \mathcal{A}_0^c = \{1, \dots, p\} \setminus \mathcal{A}_0.$$

Assumption 1 (Sampling design and graph). Locations u_i are i.i.d. with a density bounded away from zero and infinity on a compact $\mathcal{U} \subset \mathbb{R}^2$. The graph is constructed either (i) by a symmetric mutual k_n -nearest-neighbor rule with k_n bounded away from zero and $k_n = O(1)$ (node-level estimation focus), or (ii) by a symmetric compact-support kernel with bandwidth bounded away from zero. With probability tending to one the graph has $O(1)$ connected components. Degrees satisfy $\max_i d_i = O_p(1)$ and $\min_i d_i \geq c > 0$ with high probability.

Intuitively, Assumption 1 requires that locations be well spread over a compact region and that the k -NN graph does not degenerate as n grows.

Remark 1. If one targets continuum limits of $\delta_j(\cdot)$, a growing-degree regime can be imposed (e.g., $k_n \rightarrow \infty$, $k_n/n \rightarrow 0$ or shrinking bandwidth), together with a suitable scaling of the Laplacian; our theory focuses on node-level estimation and keeps $d_i = O_p(1)$.

Assumption 2 (Covariates and Gram matrices). Covariates are uniformly bounded almost surely, $\|\mathbf{Z}_i\|_\infty \leq C_Z$ and $\|\mathbf{X}_i\|_\infty \leq C_X$. The Gram matrices $n^{-1}\mathbf{Z}^\top\mathbf{Z}$ and $n^{-1}\mathbf{X}^\top\mathbf{X}$ converge to positive definite limits. For each j , $\mathbf{X}_{\odot j}$ is not identically zero on any component.

Assumption 3 (Errors and local density). Let $\varepsilon_i = Y_i - Q_\tau(Y_i | \mathbf{Z}_i, \mathbf{X}_i, u_i)$. Then $\Pr(\varepsilon_i \leq 0 | \mathbf{Z}_i, \mathbf{X}_i, u_i) = \tau$ a.s., the conditional density $f_{\varepsilon|Z,X,U}(t)$ exists and is continuous near $t = 0$, and $0 < c_f \leq f_{\varepsilon|Z,X,U}(0) \leq C_f < \infty$ a.s.

Assumption 4 (Smoothness of local deviations). For each $j \in \mathcal{A}_0^c$, the true deviation vector satisfies $\boldsymbol{\delta}_{j0}^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_{j0} \leq C_j$ with constants $C_j < \infty$ not depending on n .

Assumption 5 (Penalty sequences and adaptive weights). The smoothness penalty obeys $\lambda_{2n} \rightarrow 0$ and $n\lambda_{2n} \rightarrow \infty$. The sparsity penalty satisfies $\lambda_{1n} \rightarrow 0$ and $\sqrt{n}\lambda_{1n} \rightarrow \infty$. Adaptive weights obey $\sup_{j \in \mathcal{A}_0^c} w_{j,n} = O_p(1)$ and $\inf_{j \in \mathcal{A}_0} w_{j,n} \geq c_w > 0$ with probability tending to one.

Assumption 6 (Beta-min separation). There exists $c_\beta > 0$ such that for all $j \in \mathcal{A}_0^c$, $\|\boldsymbol{\delta}_{j0}\|_2 \geq c_\beta \lambda_{1n} w_{j,n}$ for all sufficiently large n with probability tending to one.

Assumption 7 (Restricted curvature for the quantile loss). Let $\mathbf{G}_i = (\mathbf{Z}_i^\top, \mathbf{X}_i^\top)^\top$ and define

$$\mathbf{M} = \mathbb{E}[f_{\varepsilon|Z,X,U}(0)\mathbf{G}_i\mathbf{G}_i^\top], \quad \mathbf{V} = \tau(1-\tau)\mathbb{E}[\mathbf{G}_i\mathbf{G}_i^\top].$$

Then \mathbf{M} is positive definite. Moreover, for any \mathbf{v} in a neighborhood of the origin,

$$n^{-1} \sum_{i=1}^n \{\rho_\tau(\varepsilon_i - \mathbf{G}_i^\top \mathbf{v}) - \rho_\tau(\varepsilon_i)\} = \frac{1}{2} \mathbf{v}^\top \mathbf{M} \mathbf{v} + o_p(\|\mathbf{v}\|_2^2),$$

uniformly over events where $\|\mathbf{v}\|_2 = O(n^{-1/2})$.

Remark 2. Our asymptotics cover fixed p . For growing p , additional group-sparsity conditions (e.g., restricted eigenvalues) are required and are left for future work.

KKT characterization.. Let $\psi_\tau(r) = \tau - \mathbf{1}\{r < 0\}$ and \hat{r}_i be the fitted residuals. At any global minimizer $(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L)$, the parametric block satisfies

$$\sum_{i=1}^n \psi_\tau(\hat{r}_i) \mathbf{G}_i = \mathbf{0}.$$

For each δ_j ,

$$\mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\hat{\mathbf{r}}) + 2\lambda_{2n} \mathbf{L}_{\text{sym}} \hat{\boldsymbol{\delta}}_j + \mathbf{D} \boldsymbol{\mu}_j \in -\lambda_{1n} w_{j,n} \partial \|\hat{\boldsymbol{\delta}}_j\|_2,$$

with centering constraints $\mathbf{1}_C^\top \mathbf{D} \hat{\boldsymbol{\delta}}_j = 0$ on every component and multipliers $\boldsymbol{\mu}_j$.

Deviation rates and selection consistency

Theorem 3 (Deviation rates and selection consistency). *Suppose Assumptions 1–6 hold. Then:*

1. *For each $j \in \mathcal{A}_0^c$,*

$$\frac{1}{n} \sum_{i=1}^n \{\hat{\delta}_j(u_i) - \delta_{0,j}(u_i)\}^2 = O_p\left(\frac{1}{n\lambda_{2n}} + \lambda_{2n} \mathcal{S}_j\right),$$

where \mathcal{S}_j depends on $\hat{\boldsymbol{\delta}}_{j0}^\top \mathbf{L}_{\text{sym}} \hat{\boldsymbol{\delta}}_{j0}$ and the spectrum of \mathbf{L}_{sym} . The bound is balanced at $\lambda_{2n} \asymp (n\mathcal{S}_j)^{-1/2}$.

2. *The adaptive group penalty correctly separates globally constant and truly varying covariates:*

$$\Pr(\hat{\boldsymbol{\delta}}_j = \mathbf{0} \text{ for all } j \in \mathcal{A}_0) \rightarrow 1, \quad \Pr(\hat{\boldsymbol{\delta}}_j \neq \mathbf{0} \text{ for all } j \in \mathcal{A}_0^c) \rightarrow 1.$$

On the event of correct selection, the deviations satisfy the rate in part (i).

Parametric limit distribution and oracle property

Theorem 4 (Asymptotic normality and oracle property). *Suppose Assumptions 1–7 hold. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}_{P0}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1} \mathbf{V} \mathbf{M}^{-1}).$$

Moreover, on $\{\hat{\mathcal{A}} = \mathcal{A}_0\}$ (which holds with probability tending to one by Theorem 3), the joint limit of $(\hat{\boldsymbol{\theta}}_P, \{\hat{\boldsymbol{\delta}}_j\}_{j \in \mathcal{A}_0^c})$ coincides with that of the oracle estimator that knows $\boldsymbol{\delta}_{j0} = \mathbf{0}$ for all $j \in \mathcal{A}_0$.

Lemma 1 (Cross-term control). *Under bounded covariates and fixed p , if $\lambda_{2n} \asymp n^{-1/2}$, then*

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n f_{\varepsilon|Z,X,U}(0) \left(\sum_{j=1}^p X_{ij} (\hat{\delta}_j(u_i) - \delta_{0,j}(u_i)) \right) \mathbf{G}_i \right\|_2 = o_p(1).$$

Standard error estimation.. Estimate the asymptotic variance using the plug-in

$$\hat{\mathbf{M}} = n^{-1} \sum_{i=1}^n \hat{f}_i(0) \mathbf{G}_i \mathbf{G}_i^\top, \quad \hat{\mathbf{V}} = \tau(1-\tau)n^{-1} \sum_{i=1}^n \mathbf{G}_i \mathbf{G}_i^\top,$$

where $\hat{f}_i(0)$ is a local estimator of $f_{\varepsilon|Z,X,U}(0)$ (e.g., kernel density on residuals) or via a smoothed objective with bandwidth $h_n \downarrow 0$ and $\sqrt{n}h_n \rightarrow \infty$. When spatial dependence is a concern, one may use spatial block bootstrap or Conley-type HAC corrections in the sandwich step.

5. Simulation Study

We conducted two complementary simulation experiments. *Scenario 1* uses a single relatively dense data set to provide a visual check that the proposed estimator can correctly recover the spatial structure of the coefficients under a variety of error distributions. *Scenario 2* is a Monte Carlo study that summarizes selection accuracy, estimation error, and predictive performance over repeated samples.

Both scenarios share the same basic design for spatial locations, covariates, and true coefficient surfaces; only the sample size, error distribution, and the way the results are summarized differ. The estimation algorithm (ADMM implementation of SS-SVCQR with two-stage adaptive group penalty and spatially blocked cross-validation) is exactly as described in Section 3 and is therefore not repeated here.

Settings. We place n monitoring sites on the unit square $\mathcal{U} = [0, 1]^2$. The locations $u_i = (u_{i1}, u_{i2})^\top$ are i.i.d. $\text{Unif}([0, 1]^2)$, mimicking an irregular but dense network of stations. At each site, we observe a global block $\mathbf{Z}_i \in \mathbb{R}^3$ and a potentially spatially varying block $\mathbf{X}_i \in \mathbb{R}^4$,

$$Z_{i0} \equiv 1, \quad Z_{i1} \sim \mathcal{N}(0, 1), \quad Z_{i2} \sim \mathcal{N}(0, 1), \quad X_{ij} \sim \mathcal{N}(0, 1), j = 1, \dots, 4,$$

independently in i and j .

The conditional τ th quantile is

$$Q_\tau(Y_i | \mathbf{Z}_i, \mathbf{X}_i, u_i) = \mathbf{Z}_i^\top \boldsymbol{\alpha}_0 + \sum_{j=1}^4 X_{ij} \beta_{0,j}(u_i), \quad \beta_{0,j}(u) = \beta_{G0,j} + \delta_{0,j}(u),$$

with global effects

$$\boldsymbol{\alpha}_0 = (3, -1.0, 1.5)^\top, \quad \boldsymbol{\beta}_{G0} = (5, 0, 2.5, 0)^\top,$$

so that X_1 and X_3 have nonzero global components, while X_2 and X_4 would be purely global if their spatial deviations vanished.

We design the deviation fields so that exactly two covariates are truly local:

$$\begin{aligned} \delta_{0,1}(u) &= A_1 \{ \sin(2\pi u_1) \cos(2\pi u_2) + c_1(u_1 - 0.5) \}, \\ \delta_{0,2}(u) &\equiv 0, \\ \delta_{0,3}(u) &= A_3 \{ (u_1 - 0.5)^2 + (u_2 - 0.5)^2 \}, \\ \delta_{0,4}(u) &\equiv 0, \end{aligned}$$

where (A_1, c_1, A_3) are fixed constants chosen so that the signal amplitude is comparable to the random noise. In the implementation we evaluate these fields at the observed sites, construct a k -nearest-neighbour graph on $\{u_i\}$, and apply the same degree-weighted centering operator as in Section 2 so that each $\delta_{0,j}(\cdot)$ has zero degree-weighted mean on every connected component. Under this centering, $\beta_{G0,j}$ is the degree-weighted spatial average of $\beta_{0,j}(\cdot)$ and $\delta_{0,j}(\cdot)$ represents pure spatial departures.

5.1. Recovery of Spatial Structure

Scenario 1 uses a single, relatively large dataset ($n = 1000$) for each error distribution and focuses on qualitative recovery of the spatial deviation fields. Responses are generated as

$$Y_i = Q_\tau(Y_i | \mathbf{Z}_i, \mathbf{X}_i, u_i) + \varepsilon_i,$$

with ε_i independent of $(\mathbf{Z}_i, \mathbf{X}_i, u_i)$.

We consider several error distributions to probe robustness:

- (a) Gaussian: $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$;

- (b) asymmetric Laplace (ALD) with parameter $\tau = 0.5$, which aligns the conditional median exactly with $Q_\tau(\cdot)$;
- (c) Student- t_3 noise;
- (d) contaminated normal: a 90%–10% mixture of $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(0, 25\sigma^2)$;
- (e) Cauchy errors;
- (f) heteroskedastic t_3 errors,

$$\varepsilon_i = \sigma(u_i)\eta_i, \quad \sigma(u_i) = 0.5 + 0.5u_{i1}, \quad \eta_i \sim t_3,$$

which induce a left–right gradient in noise scale.

For the contaminated normal and Cauchy cases we additionally flag as *outliers* those observations with $|\varepsilon_i|$ exceeding a high cut-off (e.g. 2.5 times the Gaussian standard deviation). These outliers are indicated as hollow circles overlaid on the *true* fields in the plots, illustrating that the proposed quantile-based estimator is robust to a small number of extreme data points.

For each error scenario, we fit SS–SVCQR to $\tau = 0.5$ using the spatially blocked cross-validation scheme of Section 3 to select (λ_1, λ_2) , and then compare the estimated deviations $\hat{\delta}_j(u_i)$ with the truth.

Figures 1–6 show, for each covariate X_j , three panels: the true deviation $\delta_{0,j}(u_i)$, the estimated deviation $\hat{\delta}_j(u_i)$, and their difference (error). All panels share a common color scale so that amplitude and sign can be compared directly across variables and scenarios. In all cases X_2 and X_4 appear essentially flat, confirming that the method does not spuriously create spatial variation for truly global covariates, while the complex patterns for X_1 and X_3 are well recovered even under heavy-tailed and heteroskedastic noise.

Figures 1–6 summarize the results in the six error scenarios. In each figure, the rows correspond to covariates X_1, \dots, X_4 , while the four columns, from left to right, show the true spatial deviation, the estimated deviation from SS–SVCQR, the pointwise error and the total effect $\beta_{G,j} + \delta_j(u)$ at the target quantile.

The first column shows the simulated *ground truth*: the deviation field $\delta_{0,j}(u)$ evaluated at the sampled locations, together with smoothed contour lines that make the underlying spatial pattern easier to discern. For X_1 this reveals an oscillatory wave-like pattern superimposed on a mild trend, while X_3 exhibits a smooth dome-shaped hotspot; by contrast, X_2 and X_4 are truly global and their deviation fields are essentially flat, apart from the overlaid outlier points in the heavy-tailed scenarios.

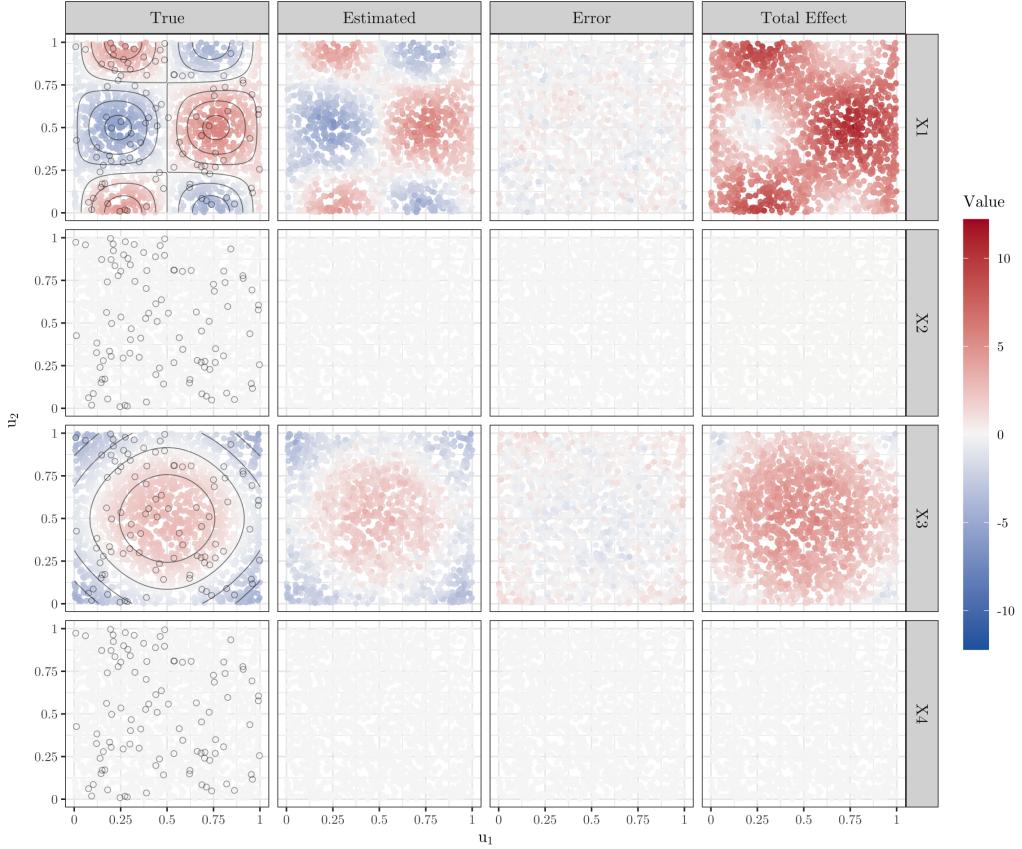


Figure 1: Scenario 1 under normal errors.

The second column contains the corresponding estimates $\hat{\delta}_j(u)$ and shows that the proposed estimator reproduces the main qualitative features of the true fields: the sign changes, level sets and apparent boundaries for X_1 and X_3 closely track those in the first column, while the panels for X_2 and X_4 remain visually featureless, indicating that spurious spatial variation is not introduced for covariates that are in fact global.

The third column displays the difference $\hat{\delta}_j(u_i) - \delta_{0,j}(u_i)$; here the colors fluctuate around zero without forming coherent spatial patches, and there is no obvious residual spatial correlation. This noise-like appearance of the error field indicates that the dominant spatial structure has been absorbed by the model and that the remaining lack-of-fit behaves as idiosyncratic noise rather than as a missing large-scale trend.

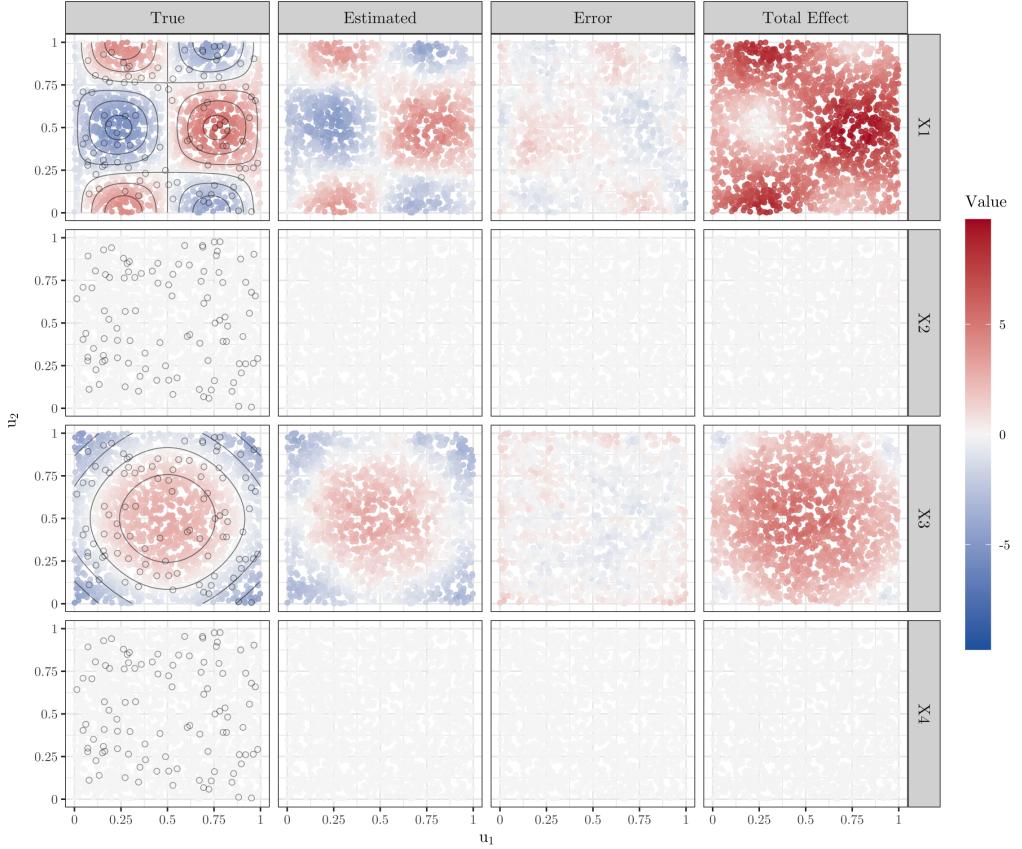


Figure 2: Scenario 1 under asymmetric Laplace errors.

Finally, the fourth column reports the total effect of each covariate in location u_i , combining the global component and the spatial deviation. For X_1 and X_3 the panels are dominated by warm colors, suggesting an overall positive association if only a single global coefficient were to be examined; however, the first column reveals sizeable regions where the local deviation is negative and partially cancels the global effect. Without explicitly separating a global level from a spatial deviation, a purely global quantile regression would report a single positive coefficient and completely miss these local reversals, potentially leading to misleading substantive conclusions.

Across all six error mechanisms, including heavy-tailed, contaminated, and heteroskedastic designs, the same qualitative picture emerges: the estimated deviation fields in the second column closely match the truth, the

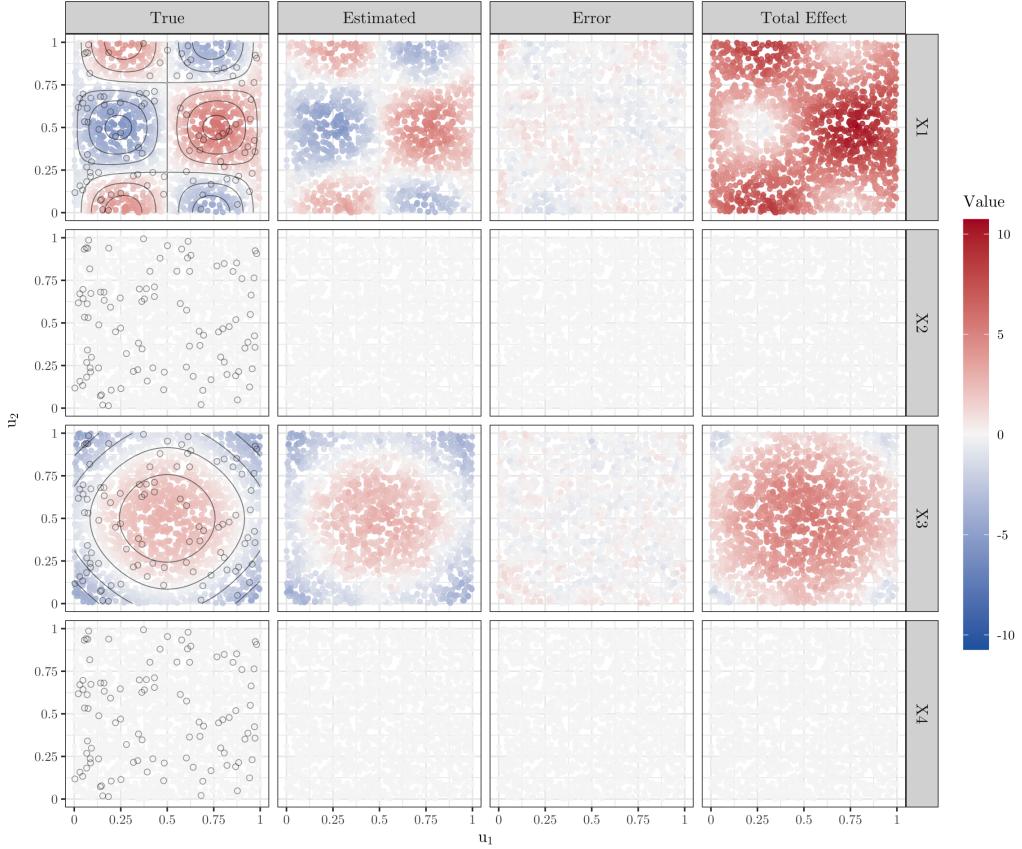


Figure 3: Scenario 1 under Student- t_3 errors.

error fields in the third column are spatially structureless, and the total-effect panels in the fourth column make clear that spatial nonstationarity can be substantial even when the overall effect is positive on average.

5.2. Monte Carlo Assessment

Scenario 2 quantifies performance over repeated samples. The spatial design, covariates and coefficient surfaces are exactly as in the common setup above, but we now generate $R = 100$ independent datasets with sample size $n = 1000$ for each error scenario of interest (e.g. Gaussian, ALD, Student- t_3 , and Cauchy). For each replicate we fit SS-SVCQR at $\tau = 0.5$ with cross-validated (λ_1, λ_2) and record:

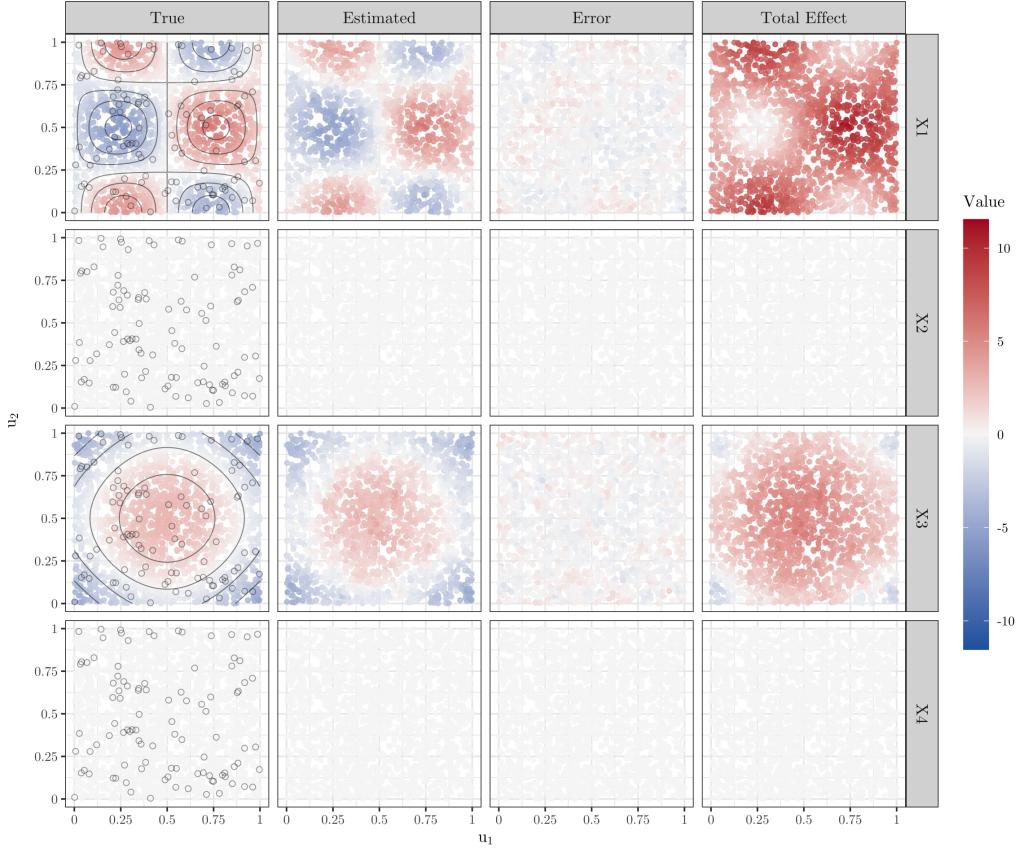


Figure 4: Scenario 1 under contaminated normal errors (hollow circles mark gross outliers).

- *Global parameter error*

$$\text{PE}_\theta = \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|_2 + \|\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_{G0}\|_2.$$

- *Deviation mean-squared error* for each covariate,

$$\text{MSE}_{\delta,j} = \frac{1}{n} \sum_{i=1}^n \{\hat{\delta}_j(u_i) - \delta_{0,j}(u_i)\}^2, \quad j = 1, \dots, 4.$$

- *Local/global classification accuracy*. We declare X_j to be “local” if

$$\sqrt{\frac{1}{n} \sum_{i=1}^n \hat{\delta}_j(u_i)^2} > \kappa$$

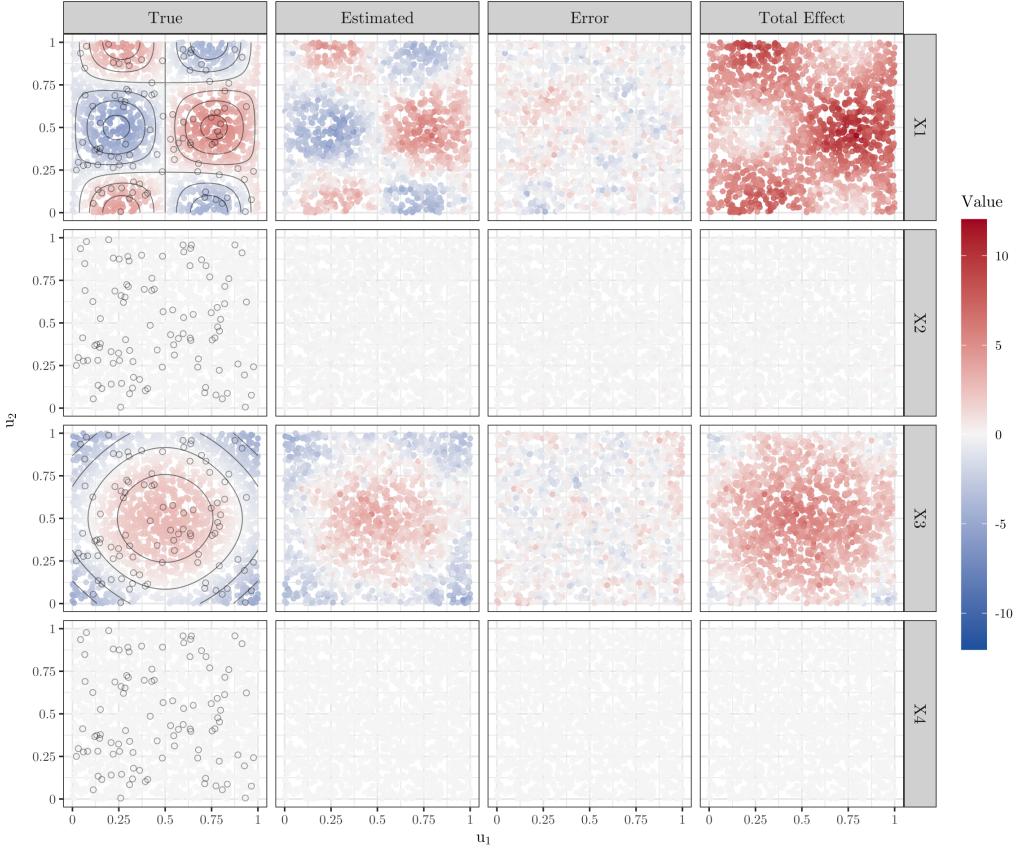


Figure 5: Scenario 1 under Cauchy errors (hollow circles mark extreme outliers).

for a small threshold κ , and “global” otherwise. Using the true set of local covariates $\{X_1, X_3\}$, we compute the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), together with sensitivity $TP/(TP + FN)$ and specificity $TN/(TN + FP)$.

- *Predictive performance* measured by the average check loss on an independent test set of size n_{test} ,

$$\text{CL}_{\text{test}} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \rho_\tau \left(Y_i^{\text{test}} - \hat{Q}_\tau(Y_i^{\text{test}} \mid \mathbf{Z}_i^{\text{test}}, \mathbf{X}_i^{\text{test}}, u_i^{\text{test}}) \right).$$

In order to objectively evaluate the performance of the proposed SS-SVCQR model, we compare it with five representative competitors. (i)

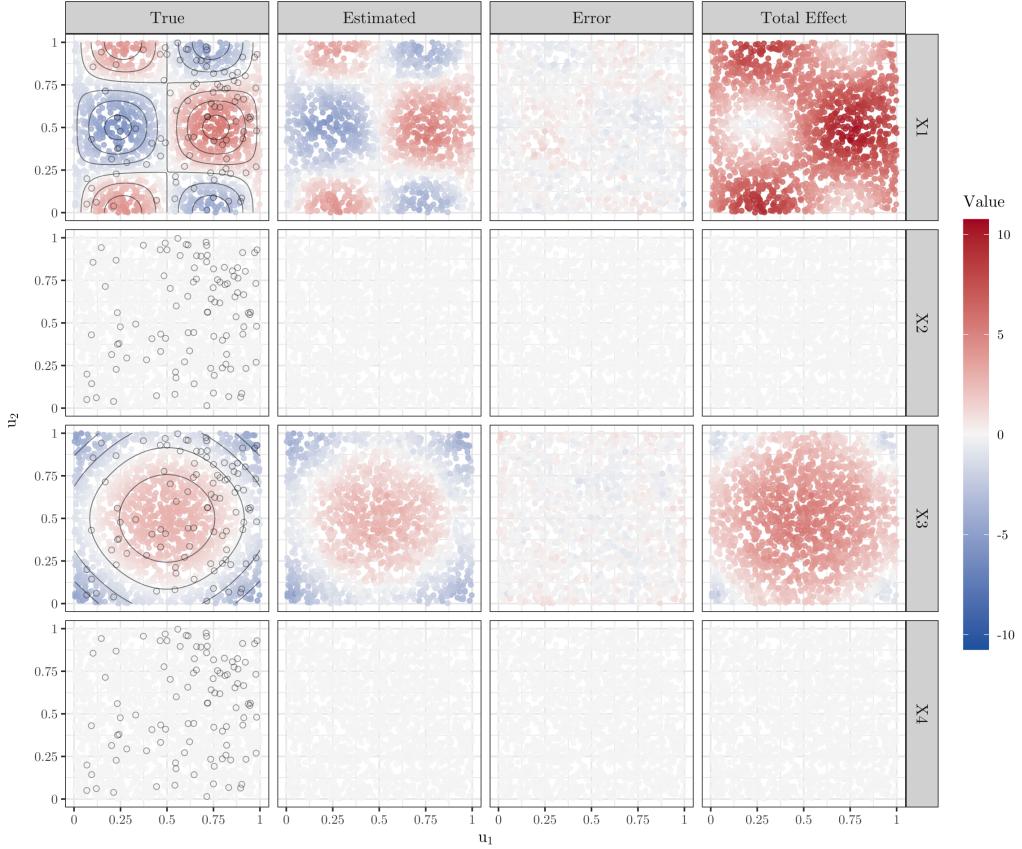


Figure 6: Scenario 1 under heteroskedastic t_3 errors with spatially varying scale.

Global quantile regression (QR) uses a purely global specification and thus ignores spatial nonstationarity[27]. (ii) Standard geographically weighted regression (GWR) treats all covariates as spatially varying[28]. (iii) Mixed GWR assumes that the local and global covariates are known a priori and fit a partially varying model; this can be regarded as an oracle benchmark for methods that must learn the local/global structure. (iv) Quantile generalized additive models (QGAM) allow flexible nonlinear effects but do not directly target spatially varying coefficients[29]. (v) The classical spatially varying coefficient (SVC) model is fitted by least squares without quantile weighting[30, 7]. These methods span the spectrum from purely global to fully spatially varying specifications and from mean to quantile regression, and they are chosen to assess three aspects: (a) accuracy of global and vary-

ing coefficient estimation, (b) ability to recover the true local versus global covariates, and (c) predictive performance under a range of error distributions. We report the detailed results of the Monte Carlo simulation in the six tables below.¹

¹In Tables 1–6, the values in parentheses represent the standard deviation corresponding to that metric.

Table 1: Monte Carlo result for quantile regression (QR)

Error	PE_θ	MSE				Sens.	Spec.	CL
		X_1	X_2	X_3	X_4			
normal	0.571 (0.157)	6.824 (0.359)	0.021 (0.026)	3.081 (0.166)	0.024 (0.036)	0 (0)	1 (0)	3.040 (0.337)
ALD	0.740 (0.215)	6.837 (0.350)	0.039 (0.060)	3.099 (0.188)	0.043 (0.062)	0 (0)	1 (0)	3.039 (0.364)
hetero	0.431 (0.117)	1.755 (0.244)	0.080 (0.027)	0.631 (0.117)	0.077 (0.028)	1 (0)	0 (0)	2.847 (0.406)
t3	0.587 (0.162)	6.814 (0.363)	0.019 (0.030)	3.089 (0.183)	0.024 (0.032)	0 (0)	1 (0)	2.998 (0.370)
contam	0.590 (0.188)	6.824 (0.368)	0.020 (0.034)	3.085 (0.177)	0.023 (0.026)	0 (0)	1 (0)	3.005 (0.317)
cauchy	0.701 (0.215)	6.839 (0.361)	0.033 (0.038)	3.100 (0.180)	0.032 (0.049)	0 (0)	1 (0)	5.803 (38.59)

Table 2: Monte Carlo result for geographically weighted regression

Error	PE_θ	MSE				Sens.	Spec.	CL
		X_1	X_2	X_3	X_4			
normal	0.471 (0.108)	2.155 (0.207)	0.064 (0.031)	0.705 (0.135)	0.065 (0.028)	1 (0)	0 (0)	2.844 (0.342)
ALD	0.615 (0.186)	2.233 (0.244)	0.152 (0.071)	0.799 (0.196)	0.134 (0.065)	1 (0)	0 (0)	2.956 (0.401)
hetero	0.483 (0.115)	2.173 (0.206)	0.071 (0.040)	0.728 (0.129)	0.070 (0.039)	1 (0)	0 (0)	2.809 (0.405)
t3	0.508 (0.122)	2.189 (0.216)	0.083 (0.046)	0.743 (0.135)	0.083 (0.044)	1 (0)	0 (0)	2.825 (0.403)
contam	0.516 (0.129)	2.179 (0.230)	0.090 (0.042)	0.730 (0.144)	0.090 (0.037)	1 (0)	0 (0)	2.877 (0.326)
cauchy	124.1 (1099.3)	9.3e5 (9.3e6)	1.2e5 (1.2e6)	8.9e5 (8.9e6)	1.5e5 (1.5e6)	1 (0)	0 (0)	40.78 (305.3)

Table 3: Monte Carlo result for mixed geographically weighted regression

Error	PE_θ	MSE				Sens.	Spec.	CL
		X_1	X_2	X_3	X_4			
normal	0.255 (0.063)	0.568 (0.117)	0.094 (0.022)	0.238 (0.053)	0.096 (0.027)	1 (0)	0 (0)	2.869 (0.384)
ALD	0.505 (0.159)	1.025 (0.251)	0.259 (0.087)	0.500 (0.154)	0.236 (0.087)	1 (0)	0 (0)	2.994 (0.404)
hetero	0.284 (0.087)	0.637 (0.147)	0.112 (0.038)	0.277 (0.075)	0.113 (0.044)	1 (0)	0 (0)	2.817 (0.411)
t3	0.337 (0.102)	0.723 (0.166)	0.142 (0.050)	0.322 (0.085)	0.146 (0.063)	1 (0)	0 (0)	2.853 (0.391)
contam	0.353 (0.098)	0.753 (0.174)	0.156 (0.056)	0.328 (0.076)	0.159 (0.044)	1 (0)	0 (0)	2.901 (0.361)
cauchy	119.5 (1062.8)	2.0e5 (2.0e6)	2.8e4 (2.7e5)	2.1e5 (2.1e6)	3.3e4 (3.3e5)	1 (0)	0 (0)	42.63 (333.9)

Table 4: Monte Carlo result for quantile generalized additive models

Error	PE_θ	MSE				Sens.	Spec.	CL
		X_1	X_2	X_3	X_4			
normal	0.248 (0.060)	1.480 (0.152)	0.019 (0.015)	0.121 (0.036)	0.020 (0.016)	1 (0)	0 (0)	2.992 (0.381)
ALD	0.427 (0.140)	1.582 (0.182)	0.062 (0.048)	0.246 (0.094)	0.060 (0.050)	1 (0)	0 (0)	3.018 (0.428)
hetero	0.230 (0.068)	1.481 (0.147)	0.016 (0.015)	0.118 (0.035)	0.017 (0.016)	1 (0)	0 (0)	2.971 (0.420)
t3	0.270 (0.076)	1.484 (0.144)	0.022 (0.019)	0.134 (0.044)	0.023 (0.020)	1 (0)	0 (0)	2.969 (0.405)
contam	0.257 (0.063)	1.508 (0.156)	0.022 (0.017)	0.129 (0.035)	0.021 (0.019)	1 (0)	0 (0)	3.001 (0.352)
cauchy	0.578 (1.299)	3.076 (11.19)	0.229 (1.675)	1.309 (9.401)	0.317 (2.566)	1 (0)	0 (0)	5.814 (38.61)

Table 5: Monte Carlo result for spatially varying coefficient model

Error	PE_θ	MSE				Sens.	Spec.	CL
		X_1	X_2	X_3	X_4			
normal	0.272 (0.066)	1.454 (0.142)	0.025 (0.020)	0.129 (0.042)	0.026 (0.020)	1 (0)	0 (0)	2.989 (0.377)
ALD	0.467 (0.151)	1.595 (0.193)	0.077 (0.052)	0.274 (0.108)	0.075 (0.056)	1 (0)	0 (0)	3.026 (0.421)
t3	0.316 (0.091)	1.488 (0.145)	0.035 (0.035)	0.171 (0.098)	0.041 (0.049)	1 (0)	0 (0)	2.962 (0.410)
contam	0.333 (0.088)	1.502 (0.146)	0.039 (0.036)	0.164 (0.048)	0.036 (0.028)	1 (0)	0 (0)	2.987 (0.365)
cauchy	114.8 (1016.2)	4.9e5 (4.9e6)	7.0e4 (7.0e5)	5.0e5 (5.0e6)	8.4e4 (8.4e5)	1 (0)	0 (0)	47.69 (371.2)

Table 6: Monte Carlo result for SS-SVCQR

Error	PE_θ	MSE				Sens.	Spec.	CL
		X_1	X_2	X_3	X_4			
normal	0.098 (0.027)	0.195 (0.014)	0 (0)	0.159 (0.010)	0 (0)	1 (0)	1 (0)	0.474 (0.009)
ALD	0.193 (0.054)	0.405 (0.035)	0 (0)	0.341 (0.038)	0 (0)	1 (0)	0.985 (0.086)	1.075 (0.023)
hetero	0.086 (0.021)	0.182 (0.012)	0 (0)	0.150 (0.009)	0 (0)	1 (0)	1 (0)	0.494 (0.012)
t3	0.113 (0.028)	0.234 (0.016)	0 (0)	0.194 (0.012)	0 (0)	1 (0)	1 (0)	0.630 (0.015)
contam	0.107 (0.032)	0.226 (0.016)	0 (0)	0.186 (0.015)	0 (0)	1 (0)	1 (0)	0.634 (0.018)
cauchy	0.156 (0.041)	0.333 (0.030)	0 (0)	0.277 (0.030)	0 (0)	1 (0)	0.990 (0.070)	6.611 (21.23)

Tables 1–6 summarize Monte Carlo means and standard deviations of the above metrics in $R = 100$ replicates in Scenario 2. For the purely global quantile regression (QR) in Table 1, the global parameter error PE_θ and the deviation MSEs are substantially larger than those of the spatial competitors whenever there is spatial nonstationarity. Moreover, QR classifies all covariates as global, leading to a sensitivity essentially equal to zero and a specificity equal to one, and its predictive performance, measured by CL_{test} , is clearly inferior once the error distribution departs from the Gaussian case.

Standard GWR and mixed GWR (Tables 2 and 3) reduce PE_θ and the deviation MSEs relative to QR, but do so by treating almost all effects as spatially varying. As a consequence, sensitivity is essentially one, while specificity is close to zero across all error distributions, indicating severe overfitting of local structure. QGAM (Table 4) shows a similar pattern: achieves smaller PE_θ and deviation MSEs than QR, but again classifies nearly all covariates as local, with sensitivity near one and specificity near zero. The classical SVC model in Table 5 behaves analogously, and under heavy-tailed Cauchy errors, its parameter error and deviation MSEs can become extremely large despite reasonably small check loss in some scenarios.

In contrast, the proposed SS–SVCQR estimator in Table 6 delivers uniformly small global parameter errors and deviation MSEs across all error settings. With the larger penalty used here, SS–SVCQR achieves sensitivity essentially equal to one and specificity close to one for normal, heteroscedastic, t_3 , and contaminated-normal errors, and only slightly below one for ALD and Cauchy errors. Thus, SS–SVCQR is the only method that simultaneously recovers both truly local covariates $\{X_1, X_3\}$ and truly global covariates $\{X_2, X_4\}$ with high precision. Its predictive check loss CL_{test} is uniformly lower than, or comparable to, that of the best competitor for light- and moderately heavy-tailed errors, and remains stable even under Cauchy errors, where several fully spatial methods suffer from dramatic loss of efficiency. These results suggest that SS–SVCQR achieves near-oracle estimation of global coefficients, reliably recovers local structure, and maintains competitive predictive performance over a wide range of error distributions in Scenario 2.

Monte Carlo averages and standard deviations of these metrics show that SS–SVCQR has high sensitivity and specificity for distinguishing local (varying) from global (constant) covariates, even under heavy-tailed and heteroskedastic errors; the deviation MSEs remain small relative to the signal variance, and the global coefficients are estimated with near-oracle precision.

Predictive performance is competitive with, or better than, fully spatially varying quantile regression and markedly improves over a purely global quantile regression when spatial nonstationarity is present.

6. Real Data

We illustrate the proposed method using the Lucas County house price data, a benchmark hedonic dataset that has been widely used in spatial econometrics and varying-coefficient modeling. The data contain 25,357 single-family home sales recorded by the Lucas County auditor between 1993 and 1998, together with structural characteristics and projected coordinates in the Ohio North State Plane system (EPSG:2834). Figure 7 maps the log-sale price for all transactions within the administrative boundary of Lucas County[31, 32].

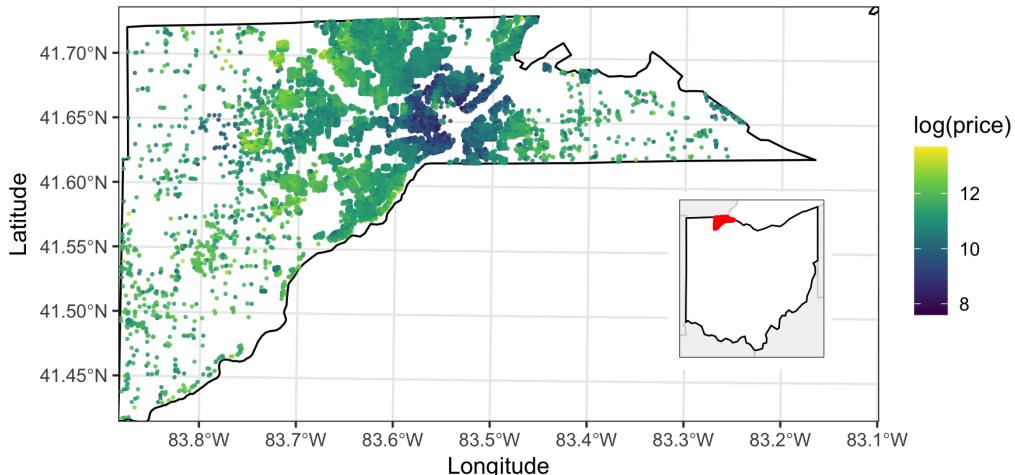


Figure 7: Spatial distribution of log house prices in Lucas County, Ohio. Points show 25,357 single-family transactions recorded by the Lucas County auditor between 1993 and 1998 (spData house dataset). Colours indicate the log sale price, with darker shades corresponding to cheaper properties and yellow-green tones to more expensive ones. The polygon outlines the administrative boundary of Lucas County. Coordinates are shown as longitude and latitude; the original data are provided in the Ohio North State Plane projection (EPSG:2834).

The map reveals pronounced spatial heterogeneity in house values. A corridor of relatively low-priced properties appears in the central urban core

along the Maumee River, whereas higher prices concentrate in the western and northwest suburbs and in a few pockets along the eastern edge of the county. The gradient is not purely radial: there are inexpensive enclaves embedded in otherwise high-priced neighborhoods and vice versa, suggesting that location interacts with local housing characteristics in a complex way rather than through a simple distance-to-CBD effect.

These patterns suggest that the marginal effects of housing attributes are unlikely to be spatially constant. For example, additional floor area may command a larger premium in high-amenity western suburbs than in older inner-city neighborhoods, and age or garage size may matter differently across locations. In addition, urban housing markets are known to be skewed and heavy-tailed, making least-squares specifications sensitive to outliers. To capture both spatially varying relationships and distributional heterogeneity, we apply our sparse-smooth spatially varying-coefficient quantile regression to model selected structural covariates with local coefficients, while keeping other controls as global effects.

Let Y_i denote the sale price of house i and $u_i \in \mathcal{U} \subset \mathbb{R}^2$ its location. Following the hedonic literature on Lucas County, we work with the log-price $Y_i^* = \log(Y_i)$, which stabilizes the right-skewed price distribution and allows the regression coefficients to be interpreted as semi-elasticities. The original coordinates are transformed to longitude-latitude and then linearly rescaled to $[0, 1]^2$ for numerical stability in the graph construction described in Section 3.

Consistent with previous hedonic studies on this data set, we focus on structural attributes of the dwelling as the main drivers of price differences: building age, interior living area, number of rooms and bedrooms, and garage size; see, for example, the specifications used in spatial autoregressive and geographically weighted models. In addition, we control for building type (number of stories, exterior wall material, garage type) and for sale year dummies that capture market-wide temporal trends.

In the notation of Section 2, we set $Y_i = Y_i^*$ and partition the covariates into a purely global block \mathbf{Z}_i and a potentially spatially varying block \mathbf{X}_i as follows. The global block collects a constant, structural-type indicators and time dummies

$$\mathbf{Z}_i = (1, \text{stories}_i, \text{wall}_i, \text{garageType}_i, \text{syear}_i, \text{baths}_i, \text{halfbaths}_i)^\top,$$

where stories_i , wall_i , garageType_i and syear_i are encoded as sets of dummy variables. These effects are interpreted as global in the sense that the impact

of being, say, a two-story brick house or selling in 1997 is assumed not to vary smoothly across the county.

The potentially spatially varying block contains continuous structural characteristics for which the hedonic literature reports marked submarket heterogeneity:

$$\mathbf{X}_i = (\text{age}_i, \log(\text{TLA}_i), \log(\text{lotsize}_i), \text{rooms}_i, \text{beds}_i, \text{garageArea}_i)^\top.$$

Here age_i is the age of the house at sale, TLA_i is the total living area (square feet), rooms_i and beds_i are the number of rooms and bedrooms, and garageArea_i is the garage floor area (with zero for houses without a garage). All components of \mathbf{X}_i are centered and scaled to unit variance before fitting so that the group penalty acts comparably across predictors. In the SS-SVCQR formulation, each coefficient is decomposed as

$$\beta_j(u) = \beta_{G,j} + \delta_j(u), \quad j = 1, \dots, p,$$

where $\beta_{G,j}$ represents the global (location-invariant) marginal effect of predictor X_{ij} on the τ th conditional quantile of log-price, and $\delta_j(u)$ captures a spatial deviation field around this global level. The adaptive group penalty in (10) decides, for each of these structural variables, whether $\delta_j(\cdot)$ should be shrunk to zero (globally constant effect) or retained as a smoothly varying coefficient.

Table 7 summarizes the variables used in the analysis and their roles in the model.

Table 7: Variables used in the Lucas County housing application and their roles in the SS–SVCQR model. SVC stands for spatially varying coefficient.

Symbol / name	Data field	Description	Role in model
Y	<code>price</code>	Sale price (USD), log-transformed	Response
u_1, u_2	coordinates	House location, rescaled to $[0, 1]^2$	Spatial index
<i>Global covariates \mathbf{Z}_i</i>			
Intercept	—	Constant term	Global effect
<code>stories</code>	<code>stories</code>	Building type (number of stories, factor)	Global effect
<code>wall</code>	<code>wall</code>	Exterior wall material (factor)	Global effect
<code>garageType</code>	<code>garage</code>	Garage type (factor)	Global effect
<code>syear</code>	<code>syear</code>	Sale year 1993–1998 (factor)	Global effect
<code>baths</code>	<code>baths</code>	Number of full bathrooms	Global effect
<code>halfbaths</code>	<code>halfbaths</code>	Number of half bathrooms	Global effect
<i>Potentially spatially varying covariates \mathbf{X}_i</i>			
X_1	<code>age</code>	House age at sale (years), standardized	SVC candidate
X_2	<code>TLA</code>	Total living area, log and standardized	SVC candidate
X_3	<code>lotsize</code>	Lot size, log and standardized	SVC candidate
X_4	<code>rooms</code>	Total number of rooms, standardized	SVC candidate
X_5	<code>beds</code>	Number of bedrooms, standardized	SVC candidate
X_6	<code>garageArea</code>	Garage area (sq ft), standardized	SVC candidate

Our choice of which regressors are allowed to vary over space is guided by both the hedonic housing literature on the Lucas County dataset and more recent studies using spatially varying coefficient (SVC) or multiscale geographically weighted regression models.

For Lucas County data, spatial econometric and semiparametric analyzes such as LeSage and Pace (2009, Chap. 7), Zhu et al. (2011), Mínguez et al. (2013), and Basile et al. (2014) consistently show that prices are mainly driven by structural characteristics: floor area, lot size, number of rooms and bedrooms, and age, together with strong spatial trend components. In contrast, design and transaction dummies (wall material, storey, garage type, year-of-sale dummies) typically enter as conventional fixed effects whose estimated coefficients are comparatively stable across space. These findings suggest that the marginal willingness to pay for size and vintage interacts more strongly with location than the effects of building style or sale year.

More broadly, recent SVC and multiscale GWR applications to hedonic pricing problems report a similar pattern. Murakami et al. (2017) and Dambon et al. (2022) document pronounced spatial variation in the coefficients of structural attributes such as floor area and age, while some categorical or regulatory variables behave almost globally. Chica-Olmo et al. (2019) and Hong et al. (2022) likewise find that structural and locational amenities often require flexible spatial effects, whereas purely institutional or temporal indicators can be treated as global without substantial loss of fit. Empirical work on housing and land prices using mixed GWR or related frameworks reaches similar conclusions, with some coefficients operating at near-global scales and others varying at very local scales (e.g., Fotheringham, Yang and Kang, 2017; Muto et al., 2023).

Motivated by this evidence, in our SS-SVCQR specification, we treat the main structural characteristics—age, total living area, lot size, number of rooms and bedrooms, and garage size—as candidate spatially varying coefficients. These variables represent the quantity and effective quality of housing services, whose marginal valuation is known to depend on local neighborhood conditions and submarkets. In contrast, the intercept, the design-related dummies (storey, wall material, garage type), the number of full and half bathrooms, and the sale-year dummies enter as global coefficients. They capture broad market conditions, building regulations, and style differences that are expected to be more homogeneous within Lucas County. Importantly, the group penalty in our estimator can still shrink a **candidate** spatially varying effect back to a global one if the data do not support substantial

Table 8: Out-of-sample performance at the median ($\tau = 0.5$) in Lucas County. The metric $CL_{0.5}$ is the average check loss on the test set; $R^2_{0.5}$ is a pseudo- R^2 relative to a constant-quantile null model.

Model	$CL_{0.5}$	$R^2_{0.5}$	Improvement over Global QR
SS-SVCQR	0.1310	0.64	+53.2%
Global QR	0.2799	0.49	—
QGAM	0.1464	0.50	+47.6%
GWR	2.4878	-7.51	strongly negative

spatial heterogeneity.

To assess out-of-sample performance, we randomly partition the 25,357 transactions into a training set (80%) and a test set (20%). The split is performed at the observation level, stratified by sale year to preserve the temporal composition of the sample across folds. All tuning parameters for SS-SVCQR (the global-local penalty λ and the graph-smoothing parameter) are selected on the training set by five-fold cross-validation, minimizing the average check loss at the target quantile. Competing models (Global QR, QGAM, and GWR) are tuned using analogous procedures: for QGAM we choose the effective degrees of freedom via REML, while for GWR we select the bandwidth by cross-validation. The reported test losses and pseudo- R^2 are computed on the held-out test set.

In the median, SS-SVCQR achieves the lowest out-of-sample check loss among all competitors (Table 8). Compared to the global QR benchmark, SS-SVCQR reduces the median check loss by approximately 53.2%. Compared to the flexible QGAM baseline, the reduction is approximately 9.5%. The GWR fit performs poorly in this setting, with substantially inflated check loss and near-zero pseudo- R^2 , reflecting its instability under local collinearity and the lack of explicit regularization.

Table 9 summarizes the decomposition of each continuous predictor into a global effect and a spatial deviation. The signs and magnitudes of the global coefficients are broadly in line with hedonic expectations and prior evidence: larger floor area and lot size are associated with higher prices, whereas older dwellings tend to be discounted.

The group-wise deviation norms reveal that bedrooms are essentially a purely global predictor: $\|\delta_{\text{beds}}\|_2$ is numerically negligible. By contrast, age, garage area, and lot size exhibit substantial spatial heterogeneity, with devi-

Table 9: Estimated global coefficients and spatial deviations for the SS–SVCQR model at $\tau = 0.5$. The column $\hat{\beta}_{G,j}$ reports the global effect, $\|\hat{\delta}_j\|_2$ is the ℓ_2 -norm of the deviation field, and $[\min \hat{\delta}_j, \max \hat{\delta}_j]$ summarizes its range over sampled locations.

Covariate X_j	$\hat{\beta}_{G,j}$	$\ \hat{\delta}_j\ _2$	$[\min \hat{\delta}_j, \max \hat{\delta}_j]$
House age (years)	−0.271	0.168	[−1.05, 1.16]
Floor area (log TLA)	0.244	0.036	[−0.27, 0.38]
Number of rooms	−0.031	0.058	[−0.57, 0.72]
Number of bedrooms	0.0067	8.4×10^{-7}	$[-7.2 \times 10^{-6}, 6.1 \times 10^{-5}]$
Garage area (sq. ft.)	−0.032	0.100	[−1.07, 0.85]
Lot size (log)	0.175	0.110	[−0.78, 0.99]

ation ranges on the order of ± 1 on the log price scale. Floor area behaves as a predominantly global covariate with only mild spatial variation, while the number of rooms shows intermediate behaviour.

Figure 8 visualizes the estimated deviation fields $\hat{\delta}_j(u)$ (left panels) and the corresponding total effects $\hat{\beta}_{G,j} + \hat{\delta}_j(u)$ (right panels).

The coefficient on dwelling age is negative on average but displays marked spatial heterogeneity. In a broad central–northern band of the county, characterized by an older housing stock, the estimated marginal effect of one additional year of age is strongly negative, indicating substantial depreciation of older units relative to newer construction. Towards several peripheral neighbourhoods, the effect becomes less negative and in some pockets slightly positive, consistent with historic or high-amenity properties commanding a premium once size and other attributes are held fixed.

The effect of floor area behaves predominantly global. The deviation surface for log TLA is comparatively smooth and of small amplitude relative to its global component, implying that an incremental increase in floor area is capitalized into house prices in a fairly homogeneous manner over Lucas County. This finding is consistent with previous hedonic studies and meta-analyzes that report a stable and strongly positive association between square footage and sale price.

The coefficients on garage area exhibit moderate spatial variation around a relatively small global effect. In most locations the marginal value of additional garage space is close to zero, with only modest positive premia in some auto-oriented neighbourhoods and slightly negative effects in others. This suggests that, conditional on overall floor area and the presence of a

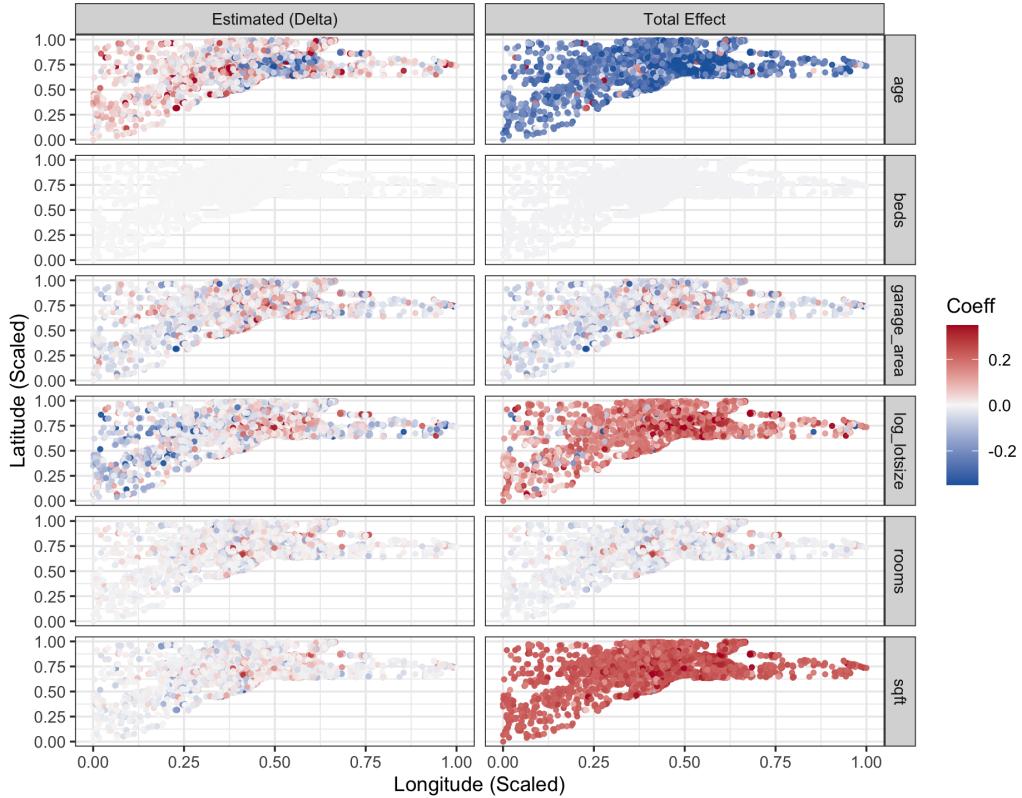


Figure 8: Estimated spatial deviations (left column) and total effects $\hat{\beta}_{G,j} + \delta_j(u)$ (right column) at $\tau = 0.5$ for the SS-SVCQR model. Each row corresponds to one covariate; colours indicate the local coefficient magnitude.

garage, extra garage square footage plays a secondary role in this market compared with indoor living space and lot size. By contrast, the lot-size premium is clearly positive across almost the entire county, with stronger effects in several lower-density suburban areas and somewhat attenuated effects in denser neighbourhoods. This pattern is in line with urban economic arguments about the trade-off between land consumption and accessibility: larger lots are valued everywhere, but their marginal contribution to price is amplified where land is relatively abundant.

Floor area (log total living area) behaves as a predominantly global covariate with a strong positive effect that varies only mildly over space. The number of rooms shows intermediate behaviour: it is positively associated with price overall, but its marginal impact fluctuates across submarkets once

floor area is controlled for. In contrast, the effect of the number of bedrooms is essentially spatially invariant, with a negligible deviation field, indicating that the valuation of an additional bedroom—conditional on total living area and rooms—can be well approximated by a single global coefficient.

Table 10: Out-of-sample quantile prediction and residual diagnostics for the Lucas County housing application.

Model	Out-of-sample performance						Diagnostics	
	CL _{0.25}	R _{0.25} ²	CL _{0.5}	R _{0.5} ²	CL _{0.75}	R _{0.75} ²	Moran's <i>I</i>	<i>p</i> -value
SS-SVCQR	0.1262	0.6132	0.1310	0.6455	0.0999	0.6471	0.126	0
Global QR	0.1326	0.4882	0.1448	0.4973	0.1026	0.5348	0.448	0
QGAM	0.1382	0.4667	0.1482	0.4856	0.1050	0.5235	0.402	0
GWR	2.1234	-7.193	2.0732	-6.196	2.0231	-8.175	0.028	0

Taken together, the results in Table 10 show that SS-SVCQR provides the best overall fit to the Lucas County data: it consistently improves out-of-sample quantile prediction relative to both global and semiparametric benchmarks, while also reducing residual spatial structure compared with purely global specifications. Combined with the coefficient patterns documented above, this evidence suggests that allowing key structural attributes to have sparse, smoothly varying effects over space is an effective and interpretable way to capture submarket heterogeneity in large housing datasets.

7. Discussion

In this paper, we have introduced a sparse–smooth spatially varying coefficient quantile regression (SS-SVCQR) framework. By integrating quantile loss with graph Laplacian regularization and group-sparse penalization, our method addresses the dual challenges of robust estimation under non-Gaussian tails and interpretable structural identification in spatial modeling.

The primary contribution of this work lies in the explicit decomposition of coefficients into global baselines and local deviations. Unlike standard Geographically Weighted Regression (GWR) [2], which often yields coefficients that fluctuate everywhere due to local sampling variability, our approach employs a group L_2 penalty to perform automated global-local model selection. This *parsimony-in-complexity* allows practitioners to distinguish between predictors that exert a universal influence across the domain and

those driven by local contextual factors. Furthermore, the use of the normalized graph Laplacian \mathbf{L}_{sym} provides a robust mechanism to handle irregular sampling designs, preventing the overfitting often observed in kernel-based methods when data density is low.

Computationally, the convex formulation of SS–SVCQR enables the use of efficient ADMM and smoothed proximal-gradient algorithms. Our theoretical analysis confirms that this computational efficiency does not come at the cost of statistical rigor, as the estimator enjoys root- n consistency and oracle selection properties under suitable conditions. From a practical perspective, these properties justify the use of SS–SVCQR as a drop-in replacement for standard QR in applications where spatial nonstationarity and heavy-tailed responses are of concern.

Despite these advantages, several limitations warrant further investigation. First, the performance of the spatial regularization relies on the construction of the adjacency graph. Although we utilized a mutual k -nearest-neighbor graph to adapt to density variations, the choice of k and the bandwidth σ remains a hyperparameter selection problem. Although cross-validation mitigates this, data-driven graph learning techniques [17] could potentially produce geometries that better reflect the underlying manifold of the spatial process. Second, our current framework focuses on a fixed quantile level τ . In applications requiring a comprehensive view of the conditional distribution, estimating multiple quantiles simultaneously with non-crossing constraints [33] would be a valuable extension. Finally, while the graph Laplacian enforces smooth variations, it penalizes abrupt changes (e.g., discontinuities across administrative boundaries or physical barriers). Incorporating edge-preserving penalties, such as Total Variation in graphs [18], could allow the model to capture piecewise constant spatial regimes more effectively.

Future work will also explore spatiotemporal extensions. By constructing a product graph over space and time, the proposed regularization framework can be naturally generalized to capture dynamic heterogeneity. Additionally, developing post-selection inference tools to quantify uncertainty for the identified local deviations remains an important open direction.

Appendix A. Proofs of Theorems

Throughout this appendix, we keep the notation of Sections 2–4. For a vector $\mathbf{v} \in \mathbb{R}^n$ we write

$$\|\mathbf{v}\|_2^2 = \sum_{i=1}^n v_i^2, \quad \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b}$$

for the Euclidean norm and inner product, and use C, c, c_0, c_1, \dots for generic finite positive constants whose value may change from line to line. The notation \preceq denotes the Löwner order on symmetric matrices. Recall the diagonal multipliers $\mathbf{X}_{\odot j} = \text{diag}(X_{1j}, \dots, X_{nj})$, the (symmetric normalized) Laplacian \mathbf{L}_{sym} in (5), the degree matrix \mathbf{D} , and the degree-weighted centering projector Proj_D in (11). The subgradient of the check loss is $\psi_\tau(r) = \tau - \mathbf{1}\{r < 0\}$ and we stack it as $\boldsymbol{\psi}_\tau(\mathbf{r}) = (\psi_\tau(r_1), \dots, \psi_\tau(r_n))^\top$.

We record two standard facts that will be used repeatedly.

Knight identity and local quadratic expansion.. For any $r, t \in \mathbb{R}$,

$$\rho_\tau(r - t) - \rho_\tau(r) = -t\psi_\tau(r) + \int_0^t \{\mathbf{1}(r \leq s) - \mathbf{1}(r \leq 0)\} ds. \quad (\text{A.1})$$

This is the well-known identity of [34]. For completeness, we briefly indicate the derivation. The function $u \mapsto \rho_\tau(u)$ is piecewise linear with a kink at 0, and its derivative exists Lebesgue-a.e. and is given by $\psi_\tau(u) = \tau - \mathbf{1}\{u < 0\}$. Integrating the derivative of $s \mapsto \rho_\tau(r - s)$ between 0 and t then yields (A.1); see also [35, Sec. 4.3].

Let $\varepsilon_i = Y_i - Q_\tau(Y_i \mid \mathbf{Z}_i, \mathbf{X}_i, u_i)$ be as in Assumption 3, and $f_{\varepsilon \mid Z, X, U}(\cdot)$ the corresponding conditional density. By (A.1), conditioning on $(\mathbf{Z}_i, \mathbf{X}_i, u_i)$ and taking expectations gives

$$\begin{aligned} \mathbb{E}\{\rho_\tau(\varepsilon_i - t) - \rho_\tau(\varepsilon_i) \mid \mathbf{Z}_i, \mathbf{X}_i, u_i\} &= -t\mathbb{E}\{\psi_\tau(\varepsilon_i) \mid \mathbf{Z}_i, \mathbf{X}_i, u_i\} \\ &\quad + \int_0^t \left(\Pr(\varepsilon_i \leq s \mid \mathbf{Z}_i, \mathbf{X}_i, u_i) - \Pr(\varepsilon_i \leq 0 \mid \mathbf{Z}_i, \mathbf{X}_i, u_i) \right) ds. \end{aligned}$$

By the defining property of conditional quantiles, $\Pr(\varepsilon_i \leq 0 \mid \mathbf{Z}_i, \mathbf{X}_i, u_i) = \tau$ and $\mathbb{E}\{\psi_\tau(\varepsilon_i) \mid \mathbf{Z}_i, \mathbf{X}_i, u_i\} = 0$. Moreover, by the continuity of the conditional density at 0 (Assumption 3)

$$\Pr(\varepsilon_i \leq s \mid \mathbf{Z}_i, \mathbf{X}_i, u_i) - \Pr(\varepsilon_i \leq 0 \mid \mathbf{Z}_i, \mathbf{X}_i, u_i) = f_{\varepsilon \mid Z, X, U}(0)s + o(s)$$

uniformly for $|s| \leq \eta$ and some $\eta > 0$. Hence, for $|t| \leq \eta$,

$$\begin{aligned} \mathbb{E} \{ \rho_\tau(\varepsilon_i - t) - \rho_\tau(\varepsilon_i) \mid \mathbf{Z}_i, \mathbf{X}_i, u_i \} &= \int_0^t \{ f_{\varepsilon|Z,X,U}(0)s + o(s) \} \, ds \\ &= \frac{1}{2} f_{\varepsilon|Z,X,U}(0)t^2 + o(t^2), \end{aligned} \quad (\text{A.2})$$

uniformly in $(\mathbf{Z}_i, \mathbf{X}_i, u_i)$. Summing over i leads to a local quadratic approximation of the empirical loss around the true residuals; cf. Assumption 7 and [35, Sec. 4.3].

Graph Poincaré inequality on the centered subspace. Let $\mathcal{S} \subset \mathbb{R}^n$ be the subspace

$$\mathcal{S} = \{ \mathbf{v} \in \mathbb{R}^n : \mathbf{1}_{\mathcal{C}}^\top \mathbf{D} \mathbf{v} = 0 \text{ for every connected component } \mathcal{C} \}.$$

By construction, the null space of \mathbf{L}_{sym} on a connected component \mathcal{C} is spanned by $\mathbf{D}^{1/2}\mathbf{1}_{\mathcal{C}}$, and therefore \mathcal{S} is orthogonal, in the \mathbf{D} -inner product, to $\ker(\mathbf{L}_{\text{sym}})$. Since \mathbf{L}_{sym} is symmetric positive semidefinite with eigenvalues in $[0, 2]$, there exists a constant $c_\lambda > 0$ (depending only on the graph sequence and bounded away from 0 under Assumption 1) such that

$$\mathbf{v}^\top \mathbf{L}_{\text{sym}} \mathbf{v} \geq c_\lambda \|\mathbf{v}\|_2^2, \quad \forall \mathbf{v} \in \mathcal{S}. \quad (\text{A.3})$$

This is a discrete Poincaré inequality; see, for instance, [22, Chap. 1] and [36, Sec. 2].

Let $(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L)$ be any global minimizer of the penalized criterion \mathcal{L}_n in Section 4. Write the true parameter as $(\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0})$ with components $\boldsymbol{\delta}_{j0}$, and define

$$q_{\tau,i}(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = \mathbf{Z}_i^\top \boldsymbol{\alpha} + \mathbf{X}_i^\top \boldsymbol{\beta}_G + \sum_{j=1}^p X_{ij} \delta_j(u_i), \quad r_i(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = Y_i - q_{\tau,i}(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L).$$

By optimality,

$$\mathcal{L}_n(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L) \leq \mathcal{L}_n(\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0}),$$

so that for any feasible $(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L)$

$$\begin{aligned} 0 &\leq \mathcal{L}_n(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) - \mathcal{L}_n(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L) \\ &= \sum_{i=1}^n \left\{ \rho_\tau(r_i(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L)) - \rho_\tau(r_i(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L)) \right\} \\ &\quad + \sum_{j=1}^p \lambda_{1n} w_{j,n} (\|\boldsymbol{\delta}_j\|_2 - \|\hat{\boldsymbol{\delta}}_j\|_2) + \sum_{j=1}^p \lambda_{2n} (\boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_j - \hat{\boldsymbol{\delta}}_j^\top \mathbf{L}_{\text{sym}} \hat{\boldsymbol{\delta}}_j). \end{aligned} \quad (\text{A.4})$$

In particular, choosing $(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = (\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0})$ yields the inequality we will work with.

In what follows we use the shorthand

$$\Delta\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0, \quad \Delta\boldsymbol{\beta}_G = \hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_{G0}, \quad \Delta\boldsymbol{\delta}_j = \hat{\boldsymbol{\delta}}_j - \boldsymbol{\delta}_{j0},$$

and denote by $\mathbf{h} = \sum_{j=1}^p \mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j \in \mathbb{R}^n$ the aggregate contribution of the deviation perturbations. Let $\mathbf{r}_0 = (\varepsilon_1, \dots, \varepsilon_n)^\top$ be the true residual vector, and set

$$t_i = q_{\tau,i}(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L) - q_{\tau,i}(\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0}) = \mathbf{Z}_i^\top \Delta\boldsymbol{\alpha} + \mathbf{X}_i^\top \Delta\boldsymbol{\beta}_G + \sum_{j=1}^p X_{ij} \Delta\boldsymbol{\delta}_j(u_i),$$

so that $r_i(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L) = \varepsilon_i - t_i$ and $r_i(\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0}) = \varepsilon_i$.

Applying Knight identity (A.1) with $r = \varepsilon_i$ and $t = t_i$ and summing over i shows that the difference of empirical losses appearing in (A.4) can be written as

$$\sum_{i=1}^n \left\{ \rho_\tau(\varepsilon_i - t_i) - \rho_\tau(\varepsilon_i) \right\} = -\langle \boldsymbol{\psi}_\tau(\boldsymbol{\varepsilon}), \mathbf{t} \rangle + \sum_{i=1}^n \int_0^{t_i} \{ \mathbf{1}(\varepsilon_i \leq s) - \mathbf{1}(\varepsilon_i \leq 0) \} ds, \quad (\text{A.5})$$

where $\mathbf{t} = (t_1, \dots, t_n)^\top$. The second term on the right-hand side admits a quadratic expansion via (A.2) and Assumption 7.

Recall the active set $\mathcal{A}_0 = \{j : \boldsymbol{\delta}_{j0} = \mathbf{0}\}$ and its complement $\mathcal{A}_0^c = \{1, \dots, p\} \setminus \mathcal{A}_0$.

Proof of Theorem 3(i). Fix a covariate index $j \in \mathcal{A}_0^c$. We control the deviation $\Delta\boldsymbol{\delta}_j$ by isolating its contribution in the basic inequality (A.4) with $(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = (\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0})$. Using (A.5) and the local quadratic approximation (A.2), we obtain

$$\begin{aligned} & \sum_{i=1}^n \left\{ \rho_\tau(\varepsilon_i - t_i) - \rho_\tau(\varepsilon_i) \right\} \\ &= -\langle \boldsymbol{\psi}_\tau(\boldsymbol{\varepsilon}), \mathbf{t} \rangle + \frac{1}{2} \sum_{i=1}^n f_{\varepsilon|Z,X,U}(0) t_i^2 + R_n(\mathbf{t}), \end{aligned} \quad (\text{A.6})$$

where $R_n(\mathbf{t}) = o_p(\|\mathbf{t}\|_2^2)$ uniformly in local neighborhoods $\{\|\mathbf{t}\|_2 \leq c\sqrt{n}\}$ by Assumption 7. Since p is fixed and the penalty sequences satisfy Assumption 5, standard arguments for penalized quantile regression (cf. [35,

Sec. 4.3]) imply $\|\Delta\boldsymbol{\alpha}\|_2 + \|\Delta\boldsymbol{\beta}_G\|_2 = O_p(n^{-1/2})$. Therefore the contribution of the parametric block to \mathbf{t} is $O_p(1)$ in ℓ_2 , and we can write

$$\mathbf{t} = \mathbf{h} + \mathbf{u}, \quad \|\mathbf{u}\|_2 = O_p(1),$$

with $\mathbf{h} = \sum_{j=1}^p \mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j$ as above. Expanding t_i^2 and using bounded covariates (Assumption 2) shows that the cross-term $\sum f(0)h_i u_i$ and the purely parametric term $\sum f(0)u_i^2$ in (A.6) are of smaller order $O_p(\|\mathbf{h}\|_2)$ and $O_p(1)$, respectively, compared to the leading quadratic term in \mathbf{h} . Absorbing these contributions into the remainder, we obtain

$$\sum_{i=1}^n \left\{ \rho_\tau(\varepsilon_i - t_i) - \rho_\tau(\varepsilon_i) \right\} = -\langle \boldsymbol{\psi}_\tau(\boldsymbol{\varepsilon}), \mathbf{h} \rangle + \frac{1}{2} \mathbf{h}^\top \mathbf{W} \mathbf{h} + o_p(\|\mathbf{h}\|_2^2), \quad (\text{A.7})$$

where $\mathbf{W} = \text{diag}\{f_{\varepsilon|Z,X,U}(0), \dots\}$ has diagonal entries uniformly bounded between c_f and C_f by Assumption 3.

Plugging this expression and $(\boldsymbol{\theta}_P, \boldsymbol{\Theta}_L) = (\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0})$ into (A.4), and then isolating the j th block, we have

$$\begin{aligned} 0 &\leq -\langle \boldsymbol{\psi}_\tau(\boldsymbol{\varepsilon}), \mathbf{h} \rangle + \frac{1}{2} \mathbf{h}^\top \mathbf{W} \mathbf{h} + o_p(\|\mathbf{h}\|_2^2) \\ &\quad + \lambda_{1n} w_{j,n} (\|\boldsymbol{\delta}_{j0}\|_2 - \|\hat{\boldsymbol{\delta}}_j\|_2) + \lambda_{2n} (\boldsymbol{\delta}_{j0}^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_{j0} - \hat{\boldsymbol{\delta}}_j^\top \mathbf{L}_{\text{sym}} \hat{\boldsymbol{\delta}}_j) \\ &\quad + \sum_{\ell \neq j} \lambda_{1n} w_{\ell,n} (\|\boldsymbol{\delta}_{\ell0}\|_2 - \|\hat{\boldsymbol{\delta}}_\ell\|_2) + \sum_{\ell \neq j} \lambda_{2n} (\boldsymbol{\delta}_{\ell0}^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_{\ell0} - \hat{\boldsymbol{\delta}}_\ell^\top \mathbf{L}_{\text{sym}} \hat{\boldsymbol{\delta}}_\ell). \end{aligned} \quad (\text{A.8})$$

Using convexity of the group norm and positive semidefiniteness of the Laplacian, the last line is nonpositive and can be dropped to obtain an inequality involving only the j th block. Writing $\mathbf{h} = \mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j + \mathbf{r}_j$ with $\mathbf{r}_j = \sum_{\ell \neq j} \mathbf{X}_{\odot \ell} \Delta\boldsymbol{\delta}_\ell$ and treating \mathbf{r}_j as fixed, the quadratic term in (A.8) yields

$$\begin{aligned} \frac{1}{2} \mathbf{h}^\top \mathbf{W} \mathbf{h} &= \frac{1}{2} \|\mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j\|_W^2 + \langle \mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j, \mathbf{W} \mathbf{r}_j \rangle + \frac{1}{2} \mathbf{r}_j^\top \mathbf{W} \mathbf{r}_j \\ &\geq \frac{1}{2} c_f \|\mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j\|_2^2 - C_f \|\mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j\|_2 \|\mathbf{r}_j\|_2 + \frac{1}{2} c_f \|\mathbf{r}_j\|_2^2. \end{aligned} \quad (\text{A.9})$$

The negative cross-term can again be absorbed into the $o_p(\|\mathbf{h}\|_2^2)$ remainder, since $\|\mathbf{r}_j\|_2 = O_p(\sqrt{n})$ and $\|\mathbf{X}_{\odot j} \Delta\boldsymbol{\delta}_j\|_2$ will be shown to be $o_p(\sqrt{n})$. Moreover, using polarization,

$$\boldsymbol{\delta}_{j0}^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_{j0} - \hat{\boldsymbol{\delta}}_j^\top \mathbf{L}_{\text{sym}} \hat{\boldsymbol{\delta}}_j = -\Delta\boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \Delta\boldsymbol{\delta}_j - 2\Delta\boldsymbol{\delta}_j^\top \mathbf{L}_{\text{sym}} \boldsymbol{\delta}_{j0}.$$

By the centering constraint, both δ_{j0} and $\hat{\delta}_j$ lie in the subspace \mathcal{S} and hence so does $\Delta\delta_j$. The Poincaré inequality (A.3) gives $\Delta\delta_j^\top \mathbf{L}_{\text{sym}} \Delta\delta_j \geq c_\lambda \|\Delta\delta_j\|_2^2$. Finally, using Cauchy–Schwarz and Assumption 4,

$$|\Delta\delta_j^\top \mathbf{L}_{\text{sym}} \delta_{j0}| \leq (\Delta\delta_j^\top \mathbf{L}_{\text{sym}} \Delta\delta_j)^{1/2} (\delta_{j0}^\top \mathbf{L}_{\text{sym}} \delta_{j0})^{1/2} \leq C \|\Delta\delta_j\|_2 \mathcal{S}_j^{1/2},$$

where we abbreviate $\mathcal{S}_j = \delta_{j0}^\top \mathbf{L}_{\text{sym}} \delta_{j0}$. Collecting these facts, the inequality (A.8) implies

$$\begin{aligned} \frac{1}{2} c_f \|\mathbf{X}_{\odot j} \Delta\delta_j\|_2^2 + \lambda_{2n} c_\lambda \|\Delta\delta_j\|_2^2 &\leq \langle \mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\boldsymbol{\varepsilon}), \Delta\delta_j \rangle + C \lambda_{2n} \mathcal{S}_j^{1/2} \|\Delta\delta_j\|_2 \\ &\quad + o_p(\|\mathbf{X}_{\odot j} \Delta\delta_j\|_2^2). \end{aligned} \tag{A.10}$$

We now bound the right-hand side. By Assumptions 2–3 and a standard moment bound,

$$\|\mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\boldsymbol{\varepsilon})\|_2 \leq C\sqrt{n}$$

with probability approaching one, so that

$$|\langle \mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\boldsymbol{\varepsilon}), \Delta\delta_j \rangle| \leq C\sqrt{n} \|\Delta\delta_j\|_2.$$

The remainder term in (A.10) is of smaller order than the leading quadratic term in $\mathbf{X}_{\odot j} \Delta\delta_j$, and can therefore be absorbed into the left-hand side. Dropping the nonnegative term $(1/2)c_f \|\mathbf{X}_{\odot j} \Delta\delta_j\|_2^2$ and dividing by $\|\Delta\delta_j\|_2$ (for $\Delta\delta_j \neq 0$) yields

$$\lambda_{2n} c_\lambda \|\Delta\delta_j\|_2 \leq C\sqrt{n} + C \lambda_{2n} \mathcal{S}_j^{1/2}.$$

Solving the quadratic inequality $au^2 \leq bu + c$ with $a = \lambda_{2n} c_\lambda$, $b = C\sqrt{n}$ and $c = C \lambda_{2n} \mathcal{S}_j^{1/2}$, we obtain

$$\|\Delta\delta_j\|_2^2 \leq C \left(\frac{n}{\lambda_{2n}^2} + \mathcal{S}_j \right).$$

Finally, using bounded covariates (Assumption 2), $\|\mathbf{X}_{\odot j} \Delta\delta_j\|_2^2$ is comparable to $\|\Delta\delta_j\|_2^2$ up to a constant factor, and hence

$$\frac{1}{n} \sum_{i=1}^n \{ \hat{\delta}_j(u_i) - \delta_{0,j}(u_i) \}^2 = \frac{\|\Delta\delta_j\|_2^2}{n} = O_p \left(\frac{1}{n \lambda_{2n}} + \lambda_{2n} \mathcal{S}_j \right),$$

which is the desired mean squared deviation bound. The two summands correspond respectively to the stochastic fluctuation (variance term) and the smoothing bias induced by the Laplacian penalty. Balancing them gives $\lambda_{2n} \asymp (n \mathcal{S}_j)^{-1/2}$. \square

Proof of Theorem 3(ii). We now show that the estimator identifies the active set \mathcal{A}_0 with probability tending to one. The argument follows the usual primal–dual witness strategy for group lasso estimators, adapted to the presence of the additional smoothness penalty; see, for example, [19, 24, 37].

Inactive groups ($j \in \mathcal{A}_0^c$). For a fixed $j \in \mathcal{A}_0^c$ we consider the KKT conditions corresponding to $\hat{\boldsymbol{\delta}}_j$. There exists a Lagrange multiplier vector $\boldsymbol{\mu}_j$ (for the centering constraint) such that

$$\mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\hat{\mathbf{r}}) + 2\lambda_{2n}\mathbf{L}_{\text{sym}}\hat{\boldsymbol{\delta}}_j + \mathbf{D}\boldsymbol{\mu}_j \in -\lambda_{1n}w_{j,n}\partial\|\hat{\boldsymbol{\delta}}_j\|_2, \quad (\text{A.11})$$

where $\hat{\mathbf{r}} = (\hat{r}_1, \dots, \hat{r}_n)^\top$ are the fitted residuals. Projecting both sides of (A.11) onto the centered subspace \mathcal{S} eliminates the multiplier term $\mathbf{D}\boldsymbol{\mu}_j$ and does not increase the norm. If we test $\hat{\boldsymbol{\delta}}_j = \mathbf{0}$, the subdifferential on the right-hand side becomes the closed Euclidean ball of radius $\lambda_{1n}w_{j,n}$. Thus a sufficient condition for $\hat{\boldsymbol{\delta}}_j = \mathbf{0}$ is

$$\|\text{Proj}_D(\mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\hat{\mathbf{r}}))\|_2 \leq \lambda_{1n}w_{j,n}.$$

Bounded covariates and Assumptions 2–3 imply $\|\mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\hat{\mathbf{r}})\|_2 = O_p(\sqrt{n})$ uniformly in j . Assumption 5 requires $\sqrt{n}\lambda_{1n} \rightarrow \infty$ and $\inf_{j \in \mathcal{A}_0^c} w_{j,n} \geq c_w > 0$ with probability tending to one, hence

$$\Pr\left(\|\text{Proj}_D(\mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\hat{\mathbf{r}}))\|_2 \leq \lambda_{1n}w_{j,n} \text{ for all } j \in \mathcal{A}_0^c\right) \rightarrow 1,$$

which implies $\Pr(\hat{\boldsymbol{\delta}}_j = \mathbf{0} \text{ for all } j \in \mathcal{A}_0^c) \rightarrow 1$.

Active groups ($j \in \mathcal{A}_0$). For $j \in \mathcal{A}_0$ consider the *restricted* problem in which we constrain $\boldsymbol{\delta}_j \equiv \mathbf{0}$ for all $j \in \mathcal{A}_0$ and optimize over $(\boldsymbol{\theta}_P, \{\boldsymbol{\delta}_j : j \in \mathcal{A}_0^c\})$. Let $(\tilde{\boldsymbol{\theta}}_P, \tilde{\boldsymbol{\Theta}}_L)$ denote its minimizer, and write $\tilde{\boldsymbol{\delta}}_j$ for the deviation blocks in the restricted solution. By essentially the same argument as in part (i), one shows that $\|\tilde{\boldsymbol{\delta}}_j - \boldsymbol{\delta}_{j0}\|_2 = O_p((n\lambda_{2n})^{-1/2} + \lambda_{2n}\mathcal{S}_j^{1/2})$ for $j \in \mathcal{A}_0^c$, and therefore $\|\tilde{\boldsymbol{\delta}}_j\|_2 \geq \|\boldsymbol{\delta}_{j0}\|_2 + o_p(1)$. Assumption 6 (groupwise beta–min condition) requires $\|\boldsymbol{\delta}_{j0}\|_2 \geq c_\beta\lambda_{1n}w_{j,n}$ for some $c_\beta > 0$ and all $j \in \mathcal{A}_0^c$. Combining these facts,

$$\Pr\left(\|\tilde{\boldsymbol{\delta}}_j\|_2 > \frac{1}{2}c_\beta\lambda_{1n}w_{j,n} \text{ for all } j \in \mathcal{A}_0^c\right) \rightarrow 1,$$

so with high probability no active group is shrunk exactly to zero in the restricted solution.

It remains to show that the restricted solution is also optimal for the *unrestricted* problem. For $j \in \mathcal{A}_0$, the KKT conditions of the restricted problem enforce

$$\|\text{Proj}_D(\mathbf{X}_{\odot j}^\top \boldsymbol{\psi}_\tau(\tilde{\mathbf{r}}))\|_2 < \lambda_{1n} w_{j,n}$$

with probability tending to one, where $\tilde{\mathbf{r}}$ are the residuals at the restricted solution. Hence one can choose subgradients $\boldsymbol{\xi}_j \in \partial\|\mathbf{0}\|_2$ with $\|\boldsymbol{\xi}_j\|_2 < 1$ such that the full KKT system for the unrestricted problem is satisfied at $(\tilde{\boldsymbol{\theta}}_P, \tilde{\boldsymbol{\Theta}}_L)$. This is the usual dual feasibility argument in the primal–dual witness construction; see [19, 37]. Therefore the restricted solution coincides with the unrestricted minimizer, and in particular $\hat{\boldsymbol{\delta}}_j = \tilde{\boldsymbol{\delta}}_j \neq \mathbf{0}$ for all $j \in \mathcal{A}_0^c$ with probability tending to one.

Combining the inactive and active cases, we conclude that

$$\Pr(\hat{\boldsymbol{\delta}}_j = \mathbf{0} \ \forall j \in \mathcal{A}_0, \ \hat{\boldsymbol{\delta}}_j \neq \mathbf{0} \ \forall j \in \mathcal{A}_0^c) \rightarrow 1,$$

and on this event the rate in part (i) holds for all $j \in \mathcal{A}_0^c$, since the preceding argument for the deviation rate did not rely on the behavior of inactive groups. \square

Proof of Lemma 1. Recall the notation $e_{j,i} = \hat{\delta}_j(u_i) - \delta_{0,j}(u_i)$ and $\mathbf{e}_j = (e_{j,1}, \dots, e_{j,n})^\top$. By Theorem 3(i),

$$\frac{1}{n} \sum_{i=1}^n e_{j,i}^2 = O_p \left(\frac{1}{n\lambda_{2n}} + \lambda_{2n} \mathcal{S}_j \right)$$

for each $j \in \mathcal{A}_0^c$. Let $\mathbf{G}_i = (\mathbf{Z}_i^\top, \mathbf{X}_i^\top)^\top$ and write

$$\mathbf{T}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n f_{\varepsilon|Z,X,U}(0) \left(\sum_{j=1}^p X_{ij} e_{j,i} \right) \mathbf{G}_i.$$

Using Cauchy–Schwarz, Assumptions 2–3, and bounded p , we get

$$\begin{aligned} \|\mathbf{T}_n\|_2 &\leq C \left(\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p X_{ij} e_{j,i} \right)^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n \|\mathbf{G}_i\|_2^2 \right)^{1/2} \\ &\leq C' \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p e_{j,i}^2 \right)^{1/2} = C' \left(\frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n e_{j,i}^2 \right)^{1/2}, \end{aligned}$$

where we used the inequality $(\sum_j a_j)^2 \leq p \sum_j a_j^2$. By the deviation bound and Assumption 4 we have

$$\frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n e_{j,i}^2 = O_p \left(\frac{1}{n\lambda_{2n}} + \lambda_{2n} \bar{\mathcal{S}} \right), \quad \bar{\mathcal{S}} = \frac{1}{p} \sum_{j=1}^p \mathcal{S}_j.$$

If $\lambda_{2n} \asymp n^{-1/2}$, as assumed in the lemma, then

$$\frac{1}{n\lambda_{2n}} + \lambda_{2n} \bar{\mathcal{S}} = O(n^{-1/2}),$$

so that $\|\mathbf{T}_n\|_2 = O_p(n^{-1/4}) = o_p(1)$. This is exactly the stated conclusion. \square

Proof of Theorem 4. We first derive the asymptotic distribution of the parametric block $\hat{\boldsymbol{\theta}}_P = (\hat{\boldsymbol{\alpha}}^\top, \hat{\boldsymbol{\beta}}_G^\top)^\top$ and then argue the oracle property.

Let \hat{r}_i be the fitted residuals and recall $\varepsilon_i = r_i(\boldsymbol{\theta}_{P0}, \boldsymbol{\Theta}_{L0})$. The KKT conditions for the parametric block read

$$\sum_{i=1}^n \psi_\tau(\hat{r}_i) \mathbf{G}_i = \mathbf{0}.$$

Write $\Delta\boldsymbol{\theta}_P = \hat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}_{P0}$. Using the representation

$$\hat{r}_i = \varepsilon_i - \mathbf{G}_i^\top \Delta\boldsymbol{\theta}_P - \sum_{j=1}^p X_{ij} \left\{ \hat{\delta}_j(u_i) - \delta_{0,j}(u_i) \right\},$$

and applying the same linearization as in (A.2) (see also [35, Sec. 4.3]), we obtain

$$\psi_\tau(\hat{r}_i) - \psi_\tau(\varepsilon_i) = -f_{\varepsilon|Z,X,U}(0) \left(\mathbf{G}_i^\top \Delta\boldsymbol{\theta}_P + \sum_{j=1}^p X_{ij} e_{j,i} \right) + \eta_i, \quad (\text{A.12})$$

where the remainder η_i satisfies $\sum_{i=1}^n \eta_i \mathbf{G}_i = o_p(\sqrt{n}\|\Delta\boldsymbol{\theta}_P\|_2) + o_p(1)$ uniformly on events where $\|\Delta\boldsymbol{\theta}_P\|_2 = O(n^{-1/2})$. Summing (A.12) over i and rearranging terms yields

$$\begin{aligned} \mathbf{0} &= \sum_{i=1}^n \psi_\tau(\hat{r}_i) \mathbf{G}_i \\ &= \sum_{i=1}^n \psi_\tau(\varepsilon_i) \mathbf{G}_i - \sum_{i=1}^n f_{\varepsilon|Z,X,U}(0) \mathbf{G}_i \mathbf{G}_i^\top \Delta\boldsymbol{\theta}_P - \sum_{i=1}^n f_{\varepsilon|Z,X,U}(0) \left(\sum_{j=1}^p X_{ij} e_{j,i} \right) \mathbf{G}_i \\ &\quad + \sum_{i=1}^n \eta_i \mathbf{G}_i. \end{aligned}$$

Dividing by \sqrt{n} and denoting

$$\mathbf{M}_n = \frac{1}{n} \sum_{i=1}^n f_{\varepsilon|Z,X,U}(0) \mathbf{G}_i \mathbf{G}_i^\top, \quad \mathbf{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(\varepsilon_i) \mathbf{G}_i,$$

we can rewrite this as

$$\sqrt{n} \mathbf{M}_n \Delta \boldsymbol{\theta}_P = \mathbf{S}_n - \frac{1}{\sqrt{n}} \sum_{i=1}^n f_{\varepsilon|Z,X,U}(0) \left(\sum_{j=1}^p X_{ij} e_{j,i} \right) \mathbf{G}_i + o_p(1).$$

Assumption 7 states that $\mathbf{M}_n \rightarrow \mathbf{M}$ in probability and that \mathbf{M} is positive definite. By Lemma 1, the second term on the right-hand side converges to zero in probability. Thus

$$\sqrt{n} \Delta \boldsymbol{\theta}_P = \mathbf{M}^{-1} \mathbf{S}_n + o_p(1).$$

By the multivariate central limit theorem and Assumption 7,

$$\mathbf{S}_n \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}), \quad \mathbf{V} = \tau(1-\tau) \text{E}(\mathbf{G}_i \mathbf{G}_i^\top),$$

and Slutsky's theorem gives

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}_{P0}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1} \mathbf{V} \mathbf{M}^{-1}),$$

which establishes the first part of the theorem.

We now argue the oracle property. Let $\hat{\mathcal{A}}$ denote the estimated set of active (truly varying) covariates, i.e. $\hat{\mathcal{A}} = \{j : \hat{\delta}_j \neq \mathbf{0}\}$. By Theorem 3(ii),

$$\Pr(\hat{\mathcal{A}} = \mathcal{A}_0) \rightarrow 1.$$

On the event $\{\hat{\mathcal{A}} = \mathcal{A}_0\}$ the minimization problem defining $(\hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\Theta}}_L)$ reduces exactly to the oracle problem in which we *a priori* set $\boldsymbol{\delta}_j \equiv \mathbf{0}$ for all $j \in \mathcal{A}_0$ and optimize over $(\boldsymbol{\theta}_P, \{\boldsymbol{\delta}_j : j \in \mathcal{A}_0^c\})$. Therefore the joint limit distribution of $(\hat{\boldsymbol{\theta}}_P, \{\hat{\delta}_j\}_{j \in \mathcal{A}_0^c})$ coincides with that of the corresponding oracle estimator. In particular, the parametric block enjoys the same asymptotic linear representation and covariance as if the true active set were known. \square

Finally, we briefly comment on variance estimation for the limiting distribution. A plug-in estimator of the asymptotic covariance matrix $\mathbf{M}^{-1} \mathbf{V} \mathbf{M}^{-1}$ is obtained by replacing \mathbf{M} and \mathbf{V} with

$$\hat{\mathbf{M}} = \frac{1}{n} \sum_{i=1}^n \hat{f}_i(0) \mathbf{G}_i \mathbf{G}_i^\top, \quad \hat{\mathbf{V}} = \tau(1-\tau) \frac{1}{n} \sum_{i=1}^n \mathbf{G}_i \mathbf{G}_i^\top,$$

where $\hat{f}_i(0)$ is either a kernel estimate of the density of ε_i at zero based on residuals, or an implicit estimate based on a smoothed check loss with bandwidth $h_n \downarrow 0$ and $\sqrt{n}h_n \rightarrow \infty$. If spatial dependence among the observations is of concern, one can further use spatial block bootstrap or heteroskedasticity-and-autocorrelation consistent (HAC) corrections in the sandwich step, but this lies beyond the main focus of the present paper.

References

- [1] C. Brunsdon, A. S. Fotheringham, M. E. Charlton, Geographically weighted regression: a method for exploring spatial nonstationarity, *Geographical Analysis* 28 (4) (1996) 281–298.
- [2] A. S. Fotheringham, C. Brunsdon, M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley, Chichester, 2002.
- [3] A. S. Fotheringham, W. Yang, W. Kang, Multiscale geographically weighted regression (mgwr), *Annals of the American Association of Geographers* 107 (6) (2017) 1247–1265. doi:10.1080/24694452.2017.1352480.
- [4] T. M. Oshan, Z. Li, W. Kang, L. J. Wolf, A. S. Fotheringham, A scalable geographically weighted regression (gwr) for investigating process spatial heterogeneity, *Computers, Environment and Urban Systems* 75 (2019) 89–103. doi:10.1016/j.compenvurbsys.2019.01.012.
- [5] C.-L. Mei, S.-Y. He, W.-J. Guo, A note on mixed geographically weighted regression, *Journal of Regional Science* 46 (4) (2006) 727–747.
- [6] G. Geniaux, D. Martinetti, A new method for dealing with spatial heterogeneity in coefficients of regressions, *Regional Science and Urban Economics* 72 (2018) 74–88, often cited as 2017 in working papers; Key kept as Geniaux2017 to match text.
- [7] A. E. Gelfand, H.-J. Kim, C. F. Sirmans, S. Banerjee, Spatial modeling with spatially varying coefficient processes, *Journal of the American Statistical Association* 98 (462) (2003) 387–396. doi:10.1198/016214503000170.

- [8] S. Banerjee, A. E. Gelfand, A. O. Finley, H. Sang, Gaussian predictive process models for large spatial data sets, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (4) (2008) 825–848.
- [9] P. H. Eilers, B. D. Marx, Splines, knots, and penalties, *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (6) (2010) 637–653.
- [10] L. Xiao, Y. Li, D. Ruppert, Fast bivariate p-splines: the sandwich smoother, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75 (3) (2013) 577–599.
- [11] R. Koenker, G. Bassett, Regression quantiles, *Econometrica* 46 (1) (1978) 33–50. doi:10.2307/1913643.
- [12] B. J. Reich, Spatiotemporal quantile regression for detecting distributional changes in environmental processes, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61 (4) (2012) 535–553. doi:10.1111/j.1467-9876.2011.01031.x.
- [13] M. Hallin, Z. Lu, K. Yu, Local linear spatial regression, *The Annals of Statistics* 32 (6) (2004) 2469–2500.
- [14] K. Lum, A. E. Gelfand, Spatial quantile multiple regression models, *Journal of the American Statistical Association* 107 (500) (2012) 1260–1273.
- [15] Z. Wang, H. Zhu, Variable selection in spatial regression models, *Journal of Agricultural, Biological, and Environmental Statistics* 14 (2) (2009) 192, distinct from Wang & Leng (2008). Cited as Wang2008 in text.
- [16] Z. Cai, X. Xu, Penalized quantile regression for varying coefficient models, *Journal of Multivariate Analysis* 122 (2013) 342–359.
- [17] D. Hallac, J. Leskovec, S. Boyd, Network lasso: Clustering and optimization in large graphs, in: *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 387–396.
- [18] Y.-X. Wang, J. Sharpnack, A. Smola, R. Tibshirani, Trend filtering on graphs, *Journal of Machine Learning Research* 17 (105) (2016) 1–41.

- [19] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (1) (2006) 49–67. doi:10.1111/j.1467-9868.2005.00532.x.
- [20] A. J. Smola, R. Kondor, Kernels and regularization on graphs, in: *Learning Theory and Kernel Machines*, Springer, 2003, pp. 144–158.
- [21] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Computation* 15 (6) (2003) 1373–1396.
- [22] F. R. K. Chung, *Spectral Graph Theory*, Vol. 92 of CBMS Regional Conference Series in Mathematics, American Mathematical Society, 1997.
- [23] H. Zou, The adaptive lasso and its oracle properties, *Journal of the American Statistical Association* 101 (476) (2006) 1418–1429.
- [24] H. Wang, C. Leng, A note on adaptive group lasso, *Computational Statistics & Data Analysis* 52 (12) (2008) 5277–5286.
- [25] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine Learning* 3 (1) (2011) 1–122.
- [26] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association* 96 (456) (2001) 1348–1360.
- [27] R. Koenker, quantreg: Quantile Regression, r package version X.Y-Z (2024).
URL <https://CRAN.R-project.org/package=quantreg>
- [28] B. Lu, I. Gollini, P. Harris, M. Charlton, C. Brunsdon, GWmodel: Geographically-Weighted Models, r package version X.Y.Z (2025).
URL <https://cran.r-project.org/package=GWmodel>
- [29] M. Fasiolo, S. N. Wood, M. Zaffran, R. Nedellec, Y. Goude, Fast calibrated additive quantile regression, *Journal of the American Statistical Association* 116 (535) (2020) 1402–1412.

- [30] T. Hastie, R. Tibshirani, Varying-coefficient models, *Journal of the Royal Statistical Society: Series B (Methodological)* 55 (4) (1993) 757–779.
- [31] R. Bivand, J. Nowosad, R. Lovelace, spData: Datasets for Spatial Analysis, r package version 2.3.4 (2025).
URL <https://jakubnowosad.com/spData/>
- [32] J. Nowosad, R. Lovelace, spDataLarge: Large datasets for spatial analysis, r package version 2.1.2 (2025).
URL <https://github.com/Nowosad/spDataLarge>
- [33] X. He, X. Pan, K. M. Tan, W.-X. Zhou, Smoothed quantile regression with large-scale inference, *Journal of Econometrics* 232 (2) (2023) 367–388. doi:10.1016/j.jeconom.2021.07.010.
- [34] K. Knight, Limiting distributions for Linear Processes with Applications to Quantile Regression, *Annals of Statistics* 26 (2) (1998) 495–527.
- [35] R. Koenker, Quantile Regression, Cambridge University Press, Cambridge, 2005. doi:10.1017/CBO9780511754098.
- [36] U. von Luxburg, A tutorial on spectral clustering, *Statistics and Computing* 17 (4) (2007) 395–416. doi:10.1007/s11222-007-9033-z.
- [37] N. Simon, J. Friedman, T. Hastie, R. Tibshirani, A sparse-group lasso, *Journal of Computational and Graphical Statistics* 22 (2) (2013) 231–245. doi:10.1080/10618600.2012.681250.