Statistica Sinica **35** (2025), 1255-1275
doi:https://doi.org/10.5705/ss.202022.0221

A HOMOGENEOUS LIKELIHOOD RATIO MEASURE FOR HIDDEN JUMP-SETS IN GENERALIZED SPATIAL REGRESSION

Pei-Sheng Lin*[1], Jun Zhu[2] and Katherine J. Curtis[2]

[1]*National Health Research Institutes and* [2]*University of Wisconsin-Madison*

Abstract: Hidden structures indicative of additional patterns relevant to the scientific inquiry are generally ignored and thus, the classical spatial regression analysis could miss important information carried by the latent variables. We develop novel methodology for uncovering some of such possible structures and patterns in spatial regression analysis. Our approach is to simultaneously model regression terms and hidden jump sets that occur abruptly across space in the presence of spatial dependence. An inequality for the homogeneity measure is derived by which we establish the consistency of jump-set selection. We devise a three-step computational algorithm based on a quasi-likelihood function and homogeneity measure to uncover patterns related to jump coefficients. Under suitable regularity conditions, we prove that the identification procedure is consistent when the hidden jump sets, covariates, and spatial correlation are incorporated into the model from the outset. The simulation study also shows sound finite-sample properties. In a case study, we examine closely county-based poverty rates in relation to industrial and racial compositions prior to the decline of manufacturing in the Upper Midwest of the U.S. Our case study reveals important socio-economic factors on poverty and additionally interesting structures and patterns not detected in classical spatial regression.

Key words and phrases: Homogeneity measure, quasi-likelihood ratio, spatial statistics.

1. Introduction

Spatial regression analysis is widely used in many scientific disciplines such as the social sciences and public health fields, for relating a response variable to explanatory variables across space while assuming spatially correlated errors (Cressie, 1993). In practice, the relation between the response and the explanatory variables is viewed as of primary interest, while accounting for spatial correlation in the error is understood to be important for a proper inference about the relation. Although sensible and popular, we believe this way of conducting regression analysis for spatial data can miss structures in the data indicative of additional patterns relevant to a scientific study. The objective of this paper is to

*Corresponding author. E-mail: pslin@nhri.edu.tw

develop and apply novel methodology that helps to uncover some of such possible structures and patterns in spatial regression analysis.

The motivating case study is from a sociological research project that studies poverty in relation to socio-economic factors. The response variable of interest is a poverty rate and the explanatory variables are the industrial structures and racial-ethnical compositions, observed at the county level in six Upper Midwestern states (Fig. 2). The traditional regression analysis, which assumes independent errors, may not be valid due to the presence of spatial dependence among counties. Spatial regression analysis assumes a regression mean as in the traditional regression analysis and models spatial correlation in the error by, for example, simultaneous autoregressive (SAR) models or conditional autoregressive (CAR) models (Cressie, 1993). Although in the traditional spatial regression, the parameters in the spatial correlation models can be estimated and inferred by likelihood-based methods, their utility in practical interpretation is generally limited. Moreover, as in any complex social and public health system, it is highly plausible that the selected explanatory variables do not fully explain the variation in the response variables and that the patterns after accounting for the explanatory variables may provide additional insight into the scientific inquiry. On the other hand, the jump sets, if identified, may reflect unknown factors and thus, are potentially valuable for eliciting additional insight, for which we will develop new models and methods. For example, in the data analysis for the Midwest data, we find additional patterns besides the explanatory variables that suggest more localized forces could also shape the poverty processes.

We propose to simultaneously model regression of the response on the explanatory variables and possible structural changes in space while accounting for spatial correlation. In particular, we consider both Gaussian and non-Gaussian responses in the exponential family of distributions. The link function features not only a regression term but also a complementary model such that the coefficient marking the change has the same absolute value but opposite signs for the true jump set and its complement set. The resulting models are no longer in the exponential family and pose a number of statistical challenges. One, the parameters in the complementary models may not be identifiable. Two, the true jump set is unknown and there are many possibilities to consider as candidate jump sets. Three, even with a correctly specified likelihood function, the computational burden can be substantial and it may be infeasible to carry out the analysis. Four, the presence of spatial correlation further complicates the estimation of parameters.

To address these challenges, the quasi-likelihood (QL) concept is adapted to estimate the model parameters and a homogeneity measure is developed from a log-quasi-likelihood (log-QL) ratio to compare candidate jump sets. Suitable QL estimating equations are derived for the regression coefficients, the jump-set coefficients, and the parameters in the spatial covariance function. For

computation, a three-step computational algorithm associated with the log-QL ratio is devised to iteratively update the parameter estimates and the jump set. In classification trees, an analogous likelihood ratio test is often used to determine a split (Zhang, 1998). In the current literature, an approximation of the log-QL ratio associated with an optimal estimating equation can be obtained by projecting the log-QL function onto a subspace spanned by estimating functions in linear forms (Li, 1993). Our innovation here is to derive new formula for choosing a suitable reference point in the QL function so that the homogeneity measure has a quadratic form, which will be shown to have considerable advantage over the linear form for finite samples such that positive deviance values are guaranteed and the computation is relatively fast. A theoretically useful inequality for the homogeneity measure is derived, by which we further establish the selection consistency in the sense that the probability of selecting the true jump set tends to one.

The proposed method can also be applied for disease mapping models in public health. To explore spatial trends for response changes over geographic regions, finding jump sets (or clusters) that attribute to localized forces besides of risk factors is important to predict disease propagation. Lin and Zhu (2020) proposed a heterogeneity measure quasi-likelihood (HM-QL) method to simultaneously analyze spatial regression and clusters, built for only Gaussian and Poisson responses. The homogeneity method we develop here applies to the broader exponential family of distributions and beyond. In addition, our homogeneity measure is based on the difference between mean responses in contrast to the heterogeneity measure based on the difference between two simple links. As a result, our approach provides a more robust metric for identifying jump sets in more flexible settings than the heterogeneity method, as can be seen in our simulation and case study.

2. Modeling and Testing

2.1. Generalized complementary models

Let $D \subset \mathbb{R}^2$ denote a continuous domain of interest. Suppose there are n observations at sampling sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in D$. At site $\boldsymbol{s}_i$, let $Y_i \equiv Y(\boldsymbol{s}_i)$ denote the response variable, let $\boldsymbol{x}_i \equiv (x_{i,1}, \ldots, x_{i,q})'$ denote a vector of q non-constant explanatory variables. As has been mentioned in the introduction, for data with a complex underlying structure, it is plausible that besides known covariates, some unknown factors exist in certain regions that elevate the intensity rates of an event. In this paper, we refer to a collection of regions whose event intensity rates jump to a level significantly higher than expected from a deterministic trend associated with $\boldsymbol{x}_i$ as a hidden jump set. Let $\Delta \subset D$ denote a hidden jump set and let $\Delta^c = D - \Delta$ denote the complement set of Δ. Also, at site $\boldsymbol{x}_i$, let ς_i denote a spatial noise from a multivariate zero-mean Gaussian distribution with variance

σ_ς^2 and a correlation model with parameters $\boldsymbol{\tau}_\varsigma$. The specific assumptions for the correlation model can be found in the Appendix.

Next, we develop an integrated model associated with $\boldsymbol{x}_i$, Δ, and a spatial noise for Y_i as follows. Let $\delta_i = I[\boldsymbol{s}_i \in \Delta]$ denote a "status variable" indicating whether site $\boldsymbol{s}_i$ belongs to the jump set Δ, where $I[\cdot]$ denotes the indicator function. Given ς_i, let $\theta_i^* = E(Y_i|\varsigma_i)$ denote a conditional mean function of Y_i. Let $g(\cdot)$ denote a link function such that

$$g(\theta_i^*) = \beta_0^* + \boldsymbol{x}_i'\boldsymbol{\beta} + \xi_\delta \delta_i + \varsigma_i, \tag{2.1}$$

where β_0 is a baseline, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)'$ is a vector of regression coefficients associated with $\boldsymbol{x}_i$, and ξ_δ denotes a jump coefficient associated with the jump set. More details about the link function can be found in Assumption 1 of the Appendix. For the spatial noise, we assume that ς_i does not involve δ_i and $\boldsymbol{x}_i$. Also, conditional on $\varsigma_1, \ldots, \varsigma_n$, $Y_1, \ldots, Y_n$ are assumed to be independent.

Since in practice, the status variable δ_i is unknown, traditional statistical inference for generalized linear mixed models can not be directly applied for the hidden model (2.1). To address this issue, we derive a complementary model associated with Y_i. Let $\theta_i = E(Y_i)$, where by the double expectation theorem, $\theta_i = E_\varsigma E(Y_i|\varsigma_i) \equiv E_\varsigma(\theta_i^*)$ and $E_\varsigma(\cdot)$ denotes an expectation over a probability space associated with the spatial noise. It thus follows from (2.1) that

$$\theta_i = E_\varsigma\{g^{-1}(\beta_0^* + \boldsymbol{x}_i'\boldsymbol{\beta} + \xi_\delta\delta_i + \varsigma_i)\}. \tag{2.2}$$

Let $\ddot{f}(u) = \partial^2 g^{-1}(u)/\partial u^2$ denote a second-order derivative of $g^{-1}(u)$ with respect to u. To find an approximation for θ_i from (2.2), we recall that ς_i does not involve $\boldsymbol{x}_i$ and δ_i. Then, on the right-hand side of (2.2), we first use a second-order Taylor series to expand $g^{-1}(\beta_0^* + \boldsymbol{x}_i'\boldsymbol{\beta} + \xi_\delta\delta_i + \varsigma_i)(\equiv \theta_i^*)$ with respect to ς_i, and then compute an approximation for $E_\varsigma(\theta_i^*)(\equiv \theta_i)$. It can be shown that $\theta_i \doteq g^{-1}\{\beta_0^* + \boldsymbol{x}_i'\boldsymbol{\beta}_i + \xi_\delta\delta_i + 0.5\ddot{f}(0)\sigma_\varsigma^2\}$. Note that the offset parameter $0.5\ddot{f}(0)\sigma_\varsigma^2$ can be adjusted by the baseline β_0. With re-scaling the intercept parameter by setting $\beta_0 = \beta_0^* + 0.5\ddot{f}(0)\sigma_\varsigma^2$, we can approximate θ_i by

$$g(\theta_i) = \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta}_i + \xi_\delta\delta_i. \tag{2.3}$$

Nevertheless, one important issue for model (2.3) is that the intercept β_0 is a combination of β_0^* from model (2.1) and spatial variation σ_ς^2 from the spatial noise, which would thus cause an identifiable problem for the intercept parameter. Details to address this issue in a computation algorithm can be seen in Section 3.1.

The model (2.3) can also be expressed as $g(\theta_i) = \beta_0 + \xi_\delta + \boldsymbol{x}_i'\boldsymbol{\beta}_i - \xi_\delta\delta_i^c$, where $\delta_i^c = I[\boldsymbol{s}_i \in \Delta^c]$ denotes a complement status variable of δ_i. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)'$, $\boldsymbol{\delta}^c = 1 - \boldsymbol{\delta} = (\delta_1^c, \ldots, \delta_n^c)'$, and

$\boldsymbol{\varsigma} = (\varsigma_1, \ldots, \varsigma_n)'$. Associated with the status vector $\boldsymbol{\delta}$, let $\boldsymbol{\theta}_\delta = (\theta_1, \ldots, \theta_n)'$. A vector form for the mean responses can be expressed as either $\boldsymbol{\theta}_\delta = g^{-1}(\beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \xi_\delta \boldsymbol{\delta})$ or $\boldsymbol{\theta}_{\delta^c} = g^{-1}(\beta_0^c + \boldsymbol{X}\boldsymbol{\beta} - \xi_\delta \boldsymbol{\delta}^c)$, where $\beta_0^c = \beta_0 + \xi_\delta$. We refer to the two forms as the complementary models for the marginal mean response vector related to $\boldsymbol{\delta}$. However, the status vector $\boldsymbol{\delta}$ for the true jump set Δ is unknown. To search for $\boldsymbol{\delta}$, let $\boldsymbol{\psi}_\alpha = (\psi_{\alpha,1}, \ldots, \psi_{\alpha,n})'$ denote a status vector for a candidate jump set Ψ_α, where $\psi_{\alpha,i} = I[\boldsymbol{s}_i \in \Psi_\alpha]$ (details on how to create candidate status vectors are discussed later). Let $\boldsymbol{\Omega} = \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N\}$ denote a collection of candidate status vectors, where N denotes the number of candidate status vectors. For simplicity, we also use $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)$ to denote $\boldsymbol{\psi}_\alpha$, unless clarity demands as in Section 3. Let $\boldsymbol{\psi}^c = 1 - \boldsymbol{\psi}$ denote a complement status vector for $\boldsymbol{\psi}$. Likewise the complementary models for the true status vector $\boldsymbol{\delta}$, for a given candidate status vector $\boldsymbol{\psi} \in \boldsymbol{\Omega}$, we model the respective mean response vectors in terms of $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^c$ by

$$\boldsymbol{\theta}_\psi = g^{-1}(\beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \xi_\psi \boldsymbol{\psi}) \quad \text{and} \quad \boldsymbol{\theta}_{\psi^c} = g^{-1}(\beta_0^c + \boldsymbol{X}\boldsymbol{\beta} + \xi_{\psi^c} \boldsymbol{\psi}^c), \tag{2.4}$$

where $\beta_0^c = \beta_0 + \xi_\psi$. Note that β_0 is fixed at the same value for all $\boldsymbol{\psi} \in \boldsymbol{\Omega}$ as that of the true model. The two forms in (2.4) are referred to as candidate complementary models associated with $\boldsymbol{\psi}$.

From the true complementary models associated with $\boldsymbol{\delta}$, we have $\boldsymbol{\theta}_\psi \neq \boldsymbol{\theta}_{\psi^c}$ in general unless $\boldsymbol{\psi} = \boldsymbol{\delta}$. Furthermore, we will show that $\boldsymbol{\theta}_\delta \equiv \boldsymbol{\theta}_\psi \equiv \boldsymbol{\theta}_{\psi^c}$ if and only if $\boldsymbol{\psi}$ is (asymptotically) equal to $\boldsymbol{\delta}$ (see details in the Appendix). Hence, one way to estimate $\boldsymbol{\delta}$ by candidate complementary models is to choose $\hat{\boldsymbol{\delta}} \in \boldsymbol{\Omega}$ such that $\boldsymbol{\theta}_{\hat{\delta}}$ and $\boldsymbol{\theta}_{\hat{\delta}^c}$, where $\hat{\boldsymbol{\delta}}^c = 1 - \hat{\boldsymbol{\delta}}$, are as close as possible. Below we develop a likelihood ratio test to measure homogeneity between $\boldsymbol{\theta}_\psi$ and $\boldsymbol{\theta}_{\psi^c}$.

2.2. A likelihood ratio test for candidate status vectors

Let $\boldsymbol{\lambda}_\psi = (\beta_0, \boldsymbol{\beta}', \xi_\psi)'$ and $\boldsymbol{\lambda}_{\psi^c} = (\beta_0^c, \boldsymbol{\beta}', \xi_{\psi^c})'$ denote the parameters of interest for the candidate complementary models (2.4). To evaluate homogeneity between the pair of generalized complementary models in (2.4), we develop a test statistic based on a QL ratio. Let $\boldsymbol{V}_{\lambda_\psi} \equiv \boldsymbol{V}(\boldsymbol{\lambda}_\psi, \boldsymbol{\tau})$ denote a covariance matrix of $\boldsymbol{Y}$ associated with (2.1), where $\boldsymbol{\tau}$ denotes a vector of covariance parameters. Note that from (2.1), we have $g\{E(\boldsymbol{Y}|\boldsymbol{\varsigma})\} = \beta_0^* + \boldsymbol{X}\boldsymbol{\beta} + \xi_\psi \boldsymbol{\psi} + \boldsymbol{\varsigma}$. The covariance matrix $\boldsymbol{V}_{\lambda_\psi}$ is therefore a function of $\boldsymbol{\lambda}_\psi$ and $\boldsymbol{\tau}$, and the parameters $\boldsymbol{\tau}$ are associated with those related to the spatial noises $\boldsymbol{\varsigma}$. More details about the covariance structure can be seen in Assumption 2. Also, let $\boldsymbol{\Lambda}_{\lambda_\psi} \equiv \boldsymbol{V}_{\lambda_\psi}^{-1}$ denote the inverse matrix of $\boldsymbol{V}_{\lambda_\psi}$. Given an observed value $\boldsymbol{y}$ of $\boldsymbol{Y}$, a QL function (McCullagh and Nelder, 1989) for correlated data,

$$\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi); \boldsymbol{\tau}\} = \int_{\boldsymbol{\theta}(\boldsymbol{\lambda}^+)}^{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)} \boldsymbol{\Lambda}(\boldsymbol{t}, \boldsymbol{\tau})(\boldsymbol{y} - \boldsymbol{t}) d\boldsymbol{t}, \tag{2.5}$$

is particularly useful for estimating the unknown parameters $\boldsymbol{\lambda}_\psi$, where $\boldsymbol{\lambda}^+$ is a given reference point.

To see whether the integration in the QL function $\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi);\boldsymbol{\tau}\}$ of (2.5) is line-independent, we first make two remarks on the covariance matrix $\boldsymbol{V}_{\lambda_\psi}$. First, by Assumption 1(e) in the Appendix, the ith component of $\boldsymbol{V}_{\lambda_\psi}$ involves only θ_i and $\boldsymbol{\tau}$. Second, recall that given $\boldsymbol{\varsigma}$, $\boldsymbol{Y}$ are independent. Since $E(Y_iY_j) = E_{\boldsymbol{\varsigma}}\{E(Y_iY_j|\boldsymbol{\varsigma})\}$, by an argument similar to the derivation for the complementary models in Section 2.1, we can conclude that $\text{cov}(Y_i, Y_j)$ involves only θ_i, θ_j, and $\boldsymbol{\tau}$. That is, the (i, j) entry of $\boldsymbol{V}_{\lambda_\psi}$ does not involve θ_k, $i \neq \jmath \neq k$. Let $\Lambda_{i,j}$ denote the (i, j)th element of $\boldsymbol{\Lambda}$. We thus have $\partial\Lambda_{i,j}/\partial\theta_k = 0$ for $i \neq j \neq k$, and therefore $\boldsymbol{V}$ satisfies the required condition for the existence of the QL function (McCullagh and Nelder, 1989).

Let $\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi);\boldsymbol{\tau}\}$ and $\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c});\boldsymbol{\tau}\}$ denote the QL functions associated with $\boldsymbol{\theta}_\psi$ and $\boldsymbol{\theta}_{\psi^c}$, respectively. A log-QL ratio between $\boldsymbol{\theta}_\psi$ and $\boldsymbol{\theta}_{\psi^c}$ is defined as

$$H^0\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi), \boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c})\} = \log\left[\frac{\sup_{\boldsymbol{\lambda}_\psi}\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi);\boldsymbol{\tau}\}}{\sup_{\boldsymbol{\lambda}_{\psi^c}}\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c});\boldsymbol{\tau}\}}\right],$$

where $\sup\Upsilon(\cdot)$ denotes the supremum function of $\Upsilon(\cdot)$. By using a first-order Taylor expansion for (2.5), the log-QL ratio can be approximated by $H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}$ as

$$H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\} = (2n)^{-1}\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi) - \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}'(\boldsymbol{\Lambda}_{\hat{\boldsymbol{\lambda}}_\psi} + \boldsymbol{\Lambda}_{\hat{\boldsymbol{\lambda}}_{\psi^c}})\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi) - \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}, \tag{2.6}$$

where $\hat{\boldsymbol{\lambda}}_\psi$ and $\hat{\boldsymbol{\lambda}}_{\psi^c}$ are QL estimates for $\boldsymbol{\lambda}_\psi$ and $\boldsymbol{\lambda}_{\psi^c}$, respectively (the QL estimates are introduced in Sec. 2.3). Details for the derivation process can be seen in the Supplementary Material. We refer to $H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}$ in (2.6) as a homogeneity measure for the pair of generalized complementary models at a given candidate status vector $\boldsymbol{\psi} \in \boldsymbol{\Omega}$.

Finally, since $\boldsymbol{\psi} \approx \boldsymbol{\delta}$ if and only if $\hat{\boldsymbol{\theta}}_\psi \approx \hat{\boldsymbol{\theta}}_{\psi^c}$, where $\hat{\boldsymbol{\theta}}_\psi = \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi)$ and $\hat{\boldsymbol{\theta}}_{\psi^c} = \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})$, we propose to estimate the true status vector $\boldsymbol{\delta}$ by minimizing the homogeneity measure (2.6)

$$\hat{\boldsymbol{\delta}} = \underset{\boldsymbol{\psi}\in\boldsymbol{\Omega}}{\text{argmin}}\, H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}. \tag{2.7}$$

In (2.7), the estimated status vector $\hat{\boldsymbol{\delta}}$ is called a homogeneous likelihood ratio (HLR) estimate for $\boldsymbol{\delta}$.

2.3. Estimating equations

To estimate the parameters in the complementary models (2.4), we develop a set of QL estimating equations. Let $\boldsymbol{Q}_1(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau}) = \partial\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi);\boldsymbol{\tau}\}/\partial\boldsymbol{\lambda}_\psi \in R^{q+2}$

denote a QL score function for $\boldsymbol{\lambda}_\psi$ by taking a derivative of $\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi);\boldsymbol{\tau}\}$ with respect to $\boldsymbol{\lambda}_\psi$. Since the QL function (2.5) is well-defined under the covariance structure $\boldsymbol{V}_{\lambda_\psi}$ as shown in Section 2.2, the QL score function has a unique root (McCullagh and Nelder, 1989) in probability as $n \to \infty$. To see this, we note from Assumption 3, $n^{-1}(\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{\theta}_\psi)'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}_\psi}(\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{\theta}_\psi)$ converges to a full-rank matrix as $n \to \infty$, where $\nabla_{\boldsymbol{a}}\boldsymbol{g} \equiv \nabla_{\boldsymbol{a}}\boldsymbol{g}(\boldsymbol{a})$ denotes a Jacobian matrix for a vector function $\boldsymbol{g}(\boldsymbol{a})$ with respect to a vector $\boldsymbol{a}$. This thus ensures that the system of equations $(\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{\theta}_\psi)'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}_\psi}\{\boldsymbol{Y} - \boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)\} = \boldsymbol{0}$ has a unique solution in probability as $n \to \infty$. More details about the relationship between the derivative of $\boldsymbol{Q}_1(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau})$ and information matrix can be seen the proof of Theorem 5 in Supplementary Material.

From (2.5), the QL score function for $\boldsymbol{\lambda}_\psi = (\beta_0, \boldsymbol{\beta}', \xi_\psi)'$ can be derived as

$$\boldsymbol{Q}_1(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau}) = (\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{\theta}_\psi)'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}_\psi}\{\boldsymbol{Y} - \boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)\}. \tag{2.8}$$

We define the first QL estimating equation as $\boldsymbol{Q}_1(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau})|_{\boldsymbol{\lambda}_\psi=\hat{\boldsymbol{\lambda}}_\psi} = \boldsymbol{0}$, where the solution $\hat{\boldsymbol{\lambda}}_\psi = (\hat{\beta}_0, \hat{\boldsymbol{\beta}}'_\psi, \hat{\xi}_\psi)'$ denotes the QL estimate for $\boldsymbol{\lambda}_\psi$. When $\boldsymbol{\theta}_\psi$ is nonlinear, a Newton-Kantovorich method (Argyros, 2008) can be applied to solve this QL estimating equation. Under Assumptions 1-3, the Newton iteration for the QL estimating equation is convergent in probability as $n \to \infty$. The related result can be seen in Theorem 5, Corollary 3, and Corollary 4, while more issues about convergence of the Newton iteration can be seen in Karimireddy et al. (2019).

Since $\beta_0^c = \beta_0 + \xi_\psi$, we estimate β_0^c by $\hat{\beta}_0^c = \hat{\beta}_0 + \hat{\xi}_\psi$. So, in the part of $\boldsymbol{\theta}_{\psi^c}$ in (2.4), now only $\boldsymbol{\beta}$ and ξ_{ψ^c} need be estimated. Let $\boldsymbol{\lambda}^*_{\psi^c} = (\hat{\beta}_0^c, \boldsymbol{\beta}', \xi_{\psi^c})'$ denote the updated parameters of $\boldsymbol{\lambda}_{\psi^c}$. Also, let $\boldsymbol{\theta}^*_{\psi^c} \equiv g^{-1}(\hat{\beta}_0^c + \boldsymbol{X}\boldsymbol{\beta} + \xi_{\psi^c}\boldsymbol{\psi}^c)$ and $\boldsymbol{\Lambda}_{\boldsymbol{\lambda}^*_{\psi^c}} \equiv \boldsymbol{\Lambda}\{(\hat{\beta}_0^c, \boldsymbol{\beta}', \xi_{\psi^c}), \boldsymbol{\tau}\}$. A QL score function for $\boldsymbol{\lambda}^*_{\psi^c}$ is given by

$$\boldsymbol{Q}_2(\boldsymbol{\lambda}^*_{\psi^c};\boldsymbol{\tau}) = (\nabla_{\boldsymbol{\lambda}^*_{\psi^c}}\boldsymbol{\theta}^*_{\psi^c})'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}^*_{\psi^c}}(\boldsymbol{Y} - \boldsymbol{\theta}^*_{\psi^c}). \tag{2.9}$$

We define the second QL estimating equation for $\boldsymbol{\beta}$ and ξ_{ψ^c} as $\boldsymbol{Q}_2(\boldsymbol{\lambda}^*_{\psi^c};\boldsymbol{\tau})|_{\boldsymbol{\lambda}^*_{\psi^c}=\hat{\boldsymbol{\lambda}}^*_{\psi^c}} = \boldsymbol{0}$, where the solution $\hat{\boldsymbol{\lambda}}^*_{\psi^c} = (\hat{\beta}_0^c, \hat{\boldsymbol{\beta}}'_{\psi^c}, \hat{\xi}_{\psi^c})'$ denotes the QL estimate. Also, $\hat{\boldsymbol{\lambda}}_{\psi^c} \equiv \hat{\boldsymbol{\lambda}}^*_{\psi^c} = (\hat{\beta}_0^c, \hat{\boldsymbol{\beta}}'_{\psi^c}, \hat{\xi}_{\psi^c})'$ can be regarded as the QL estimate for $\boldsymbol{\lambda}_{\psi^c}$.

Finally, we use a variogram to estimate the covariance parameters $\boldsymbol{\tau}$. Let $\hat{\boldsymbol{\theta}}_\psi = g^{-1}\{\hat{\beta}_0 + \boldsymbol{X}\hat{\boldsymbol{\beta}}_\psi + \hat{\xi}_\psi\boldsymbol{\psi}\}$ denote an estimate of $\boldsymbol{\theta}_\psi$ and let $\boldsymbol{\epsilon}_\psi = \boldsymbol{Y} - \hat{\boldsymbol{\theta}}_\psi = (\epsilon(\boldsymbol{s}_1), \ldots, \epsilon(\boldsymbol{s}_n))'$ denote a vector of residuals, which contain information about the spatial errors. For a spatial random field, a variogram $\gamma(h) = \text{var}\{\epsilon(\boldsymbol{s}_i) - \epsilon(\boldsymbol{s}_j)\}$, where $h = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ denotes the Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, is commonly used to quantify spatial correlation. The variogram can be estimated by an empirical variogram $\hat{\gamma}(h) = \sum_{(\boldsymbol{s}_i,\boldsymbol{s}_j)\in\mathcal{E}_h)}\{\epsilon(\boldsymbol{s}_i) - \epsilon(\boldsymbol{s}_j)\}^2/|\mathcal{E}_h|$, where $|\cdot|$ denotes the cardinality of a set and $\mathcal{E}_h = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : h - c_0 < \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq h + c_0\}$ for some $c_0 > 0$ (Cressie, 1993). For a given variogram model $\gamma_{\boldsymbol{\tau}}(h)$, the parameters

$\boldsymbol{\tau}$ can be estimated by minimizing a weighted sum of squares:

$$\hat{\boldsymbol{\tau}} = \underset{\boldsymbol{\tau}}{\operatorname{argmin}} \sum_{h} |\mathcal{E}_h| \left\{ \frac{\hat{\gamma}(h)}{\gamma_{\boldsymbol{\tau}}(h)} - 1 \right\}^2. \tag{2.10}$$

Details about the computational procedure are given in Section 3.1.

3. Jump Set Identification

3.1. Single jump set identification

Recall that $\boldsymbol{\Omega} = \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N\}$ is the collection of candidate status vectors, and we aim to choose $\hat{\boldsymbol{\delta}} \in \boldsymbol{\Omega}$ such that $H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\hat{\delta}}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\hat{\delta}^c})\}$ is minimized. In practice, there are two issues for selecting suitable $\hat{\boldsymbol{\delta}}$ from $\boldsymbol{\Omega}$: (i) the number of candidate status vectors in $\boldsymbol{\Omega}$ may be large, and (ii) most of the candidate jump sets are misspecified. Recall that in the approximation process for θ_i from (2.2) to (2.3), the intercept β_0 is adjusted by an initial intercept β_0^* of (2.1) and spatial variation σ_ς^2. Since the variogram estimation (2.10) is sensitive to misspecified models, and the intercept estimation could also be affected by the offset parameter related to spatial noises, our estimation method should be calibrated to ensure that every candidate model has a same baseline. To maintain the intercept and covariance estimates to be the same for all candidate models, we propose a three-step computational algorithm, associated with a test to address the multiple testing issue, to iteratively update the estimates of parameters and status vectors.

To develop the test for parameters, first note that it may happen that $H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\} = 0$ but there is no jump set. So, before using (2.7) to select the status vector, we use the following procedure to test significance of $\hat{\xi}_\psi$. Specifically, let $\mathcal{H}_\psi$ denote the null hypothesis that $\xi_\psi \le 0$ for $\psi \in \boldsymbol{\Omega}$. A QL Z-test statistic, $Z_\psi = \hat{\xi}_\psi / \sigma_\psi$, is then used to test $\mathcal{H}_\psi$, where $\sigma_\psi^2 = \operatorname{var}(\hat{\xi}_\psi)$. To compute σ_ψ, we let $\boldsymbol{U}_\psi = \{(\nabla_{\xi_\psi}\boldsymbol{\theta}_\psi)'\boldsymbol{\Lambda}_{\lambda_\psi}(\nabla_{\xi_\psi}\boldsymbol{\theta}_\psi)\}^{-1}(\nabla_{\xi_\psi}\boldsymbol{\theta}_\psi)'\boldsymbol{\Lambda}_{\lambda_\psi}$. Under $\mathcal{H}_\psi$ and the conditions given in Theorem 4 of the Appendix, we can show that, as $n \to \infty$,

$$\sigma_\psi^2 = \boldsymbol{U}_\psi \boldsymbol{V}_{\lambda_\psi} \boldsymbol{U}_\psi' + o_p(1). \tag{3.1}$$

However, in practice, $\boldsymbol{V}_{\lambda_\psi}$ is unknown. We thus estimate $\boldsymbol{U}_\psi$ by

$$\hat{\boldsymbol{U}}_\psi = \{(\nabla_{\xi_\psi}\hat{\boldsymbol{\theta}}_\psi)'\hat{\boldsymbol{\Lambda}}_{\lambda_\psi}(\nabla_{\xi_\psi}\hat{\boldsymbol{\theta}}_\psi)\}^{-1}(\nabla_{\xi_\psi}\hat{\boldsymbol{\theta}}_\psi)'\hat{\boldsymbol{\Lambda}}_{\lambda_\psi},$$

where $\nabla_{\xi_\psi}\hat{\boldsymbol{\theta}}_\psi = \nabla_{\xi_\psi}\boldsymbol{\theta}_\psi|_{\boldsymbol{\lambda}_\psi = \hat{\boldsymbol{\lambda}}_\psi}$ and $\hat{\boldsymbol{\Lambda}}_{\lambda_\psi} = \{\boldsymbol{V}(\hat{\boldsymbol{\lambda}}_\psi; \hat{\boldsymbol{\tau}})\}^{-1}$. A sandwich estimate for σ_ψ^2 is given by $\hat{\sigma}_\psi^2 = \hat{\boldsymbol{U}}_\psi \tilde{\boldsymbol{V}}_{\lambda_\psi} \hat{\boldsymbol{U}}_\psi'$, where $\tilde{\boldsymbol{V}}_{\lambda_\psi}$ is an estimate of $\boldsymbol{V}_{\lambda_\psi}$ by a sample covariance matrix of $\boldsymbol{Y}$. Let φ denote a desired significance level for $\mathcal{H}_\psi$. We thus reject $\mathcal{H}_\psi$ if $P(Z \ge Z_\psi) \le 1 - (1-\varphi)^{1/N}$, where Z denotes the standard normal random variable.

In the following algorithm, for each iteration m, $m = 1, 2, \ldots$, let $\hat{\beta}_0^{(m)}$, $\hat{\boldsymbol{\tau}}^{(m)}$, and $\hat{\boldsymbol{\delta}}^{(m)}$ denote the estimates for the intercept β_0, covariance parameters $\boldsymbol{\tau}$, and status vector $\boldsymbol{\delta}$, respectively. In the initial step of the algorithm (i.e., $m = 0$), assume that $\hat{\boldsymbol{\tau}}^{(0)} \equiv \mathbf{0}$. Then, we fit the complementary model (2.4) for each $\boldsymbol{\psi} \in \boldsymbol{\Omega}$ with the intercept estimate being fixed at $\hat{\beta}_0^{(0)} = \sum Y_i/n$. We use (2.8) and (2.9) to obtain initial estimates $\hat{\boldsymbol{\lambda}}_{\psi}^{(0)} = (\hat{\beta}_0^{(0)}, \hat{\boldsymbol{\beta}}_{\psi}^{(0)'}, \hat{\xi}_{\psi}^{(0)})'$ for $\boldsymbol{\lambda}_{\psi}$ and $\hat{\boldsymbol{\lambda}}_{\psi^c}^{*(0)} = (\hat{\beta}_0^{(0)} + \hat{\xi}_{\psi^c}^{(0)}, \hat{\boldsymbol{\beta}}_{\psi^c}^{(0)'}, \hat{\xi}_{\psi^c}^{(0)})'$ for $\boldsymbol{\lambda}_{\psi^c}^{*}$. Import $\hat{\boldsymbol{\lambda}}_{\psi}^{(0)}$ and $\hat{\boldsymbol{\lambda}}_{\psi^c}^{(0)} \equiv \hat{\boldsymbol{\lambda}}_{\psi^c}^{*(0)}$ into (2.7) to get an initial estimate $\hat{\boldsymbol{\delta}}^{(0)}$ for the status vector. Then, for iteration $m = 1, \ldots$, conduct the following Algorithm 1.

Algorithm 1 Three-Step Algorithm.

Step 1. Update the intercept and covariance estimates

(i) Given $\hat{\boldsymbol{\delta}}^{(m-1)}$ and $\hat{\boldsymbol{\tau}}^{(m-1)}$, use $\boldsymbol{Q}_1\{\boldsymbol{\lambda}_{\hat{\delta}^{(m-1)}}; \hat{\boldsymbol{\tau}}^{(m-1)}\} = \mathbf{0}$ in (2.8) to get QL estimates $\hat{\boldsymbol{\lambda}}_{\hat{\delta}^{(m-1)}} = \big(\hat{\beta}_{0,\hat{\delta}^{(m-1)}}, \hat{\boldsymbol{\beta}}'_{\hat{\delta}^{(m-1)}}, \hat{\xi}_{\hat{\delta}^{(m-1)}}\big)'$. Update the intercept estimate to $\hat{\beta}_0^{(m)} = \hat{\beta}_{0,\hat{\delta}^{(m-1)}}$.

(ii) Let $\boldsymbol{\epsilon}_{\hat{\delta}^{(m-1)}} = \boldsymbol{Y} - \hat{\boldsymbol{\theta}}_{\hat{\delta}^{(m-1)}}$. Update $\hat{\boldsymbol{\tau}}^{(m-1)}$ to $\hat{\boldsymbol{\tau}}^{(m)}$ by (2.10).

Step 2. Update estimates of jump coefficients for candidate models

(i) Let $\boldsymbol{\lambda}_{\psi}^{*} = (\hat{\beta}_0^{(m)}, \boldsymbol{\beta}', \xi_{\psi})'$. Use $\boldsymbol{Q}_1\{\boldsymbol{\lambda}_{\psi}^{*}; \hat{\boldsymbol{\tau}}^{(m)}\} = \mathbf{0}$ in (2.8) to get QL estimates $\hat{\boldsymbol{\lambda}}_{\psi}^{*(m)} = (\hat{\beta}_0^{(m)}, \hat{\boldsymbol{\beta}}_{\psi}^{(m)'}, \hat{\xi}_{\psi}^{(m)})'$.

(ii) Use $\boldsymbol{Q}_2\{\boldsymbol{\lambda}_{\psi^c}^{*}; \hat{\boldsymbol{\tau}}^{(m)}\} = \mathbf{0}$ in (2.9) to get QL estimates $\hat{\boldsymbol{\lambda}}_{\psi^c}^{*(m)} = (\hat{\beta}_0^{(m)} + \hat{\xi}_{\psi}^{(m)}, \hat{\boldsymbol{\beta}}_{\psi^c}^{(m)'}, \hat{\xi}_{\psi^c}^{(m)})'$.

Step 3. Update an estimate for the status vector

(i) Test significance of $\hat{\xi}_{\psi}^{(m)}$ by the QL Z-test.

(ii) For $\boldsymbol{\psi}$ with significant $\hat{\xi}_{\psi}^{(m)}$, import $\hat{\boldsymbol{\lambda}}_{\psi}^{(m)} \equiv \hat{\boldsymbol{\lambda}}_{\psi}^{*(m)}$ and $\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)} \equiv \hat{\boldsymbol{\lambda}}_{\psi^c}^{*(m)}$ from Step 2 into (2.7) to update $\hat{\boldsymbol{\delta}}^{(m-1)}$ to $\hat{\boldsymbol{\delta}}^{(m)}$.

In Algorithm 1, we stop the iteration if $\hat{\boldsymbol{\delta}}^{(m)} = \hat{\boldsymbol{\delta}}^{(m-1)}$. Otherwise, let $m = m + 1$ and repeat Steps 1–3. Since the homogeneity measure (2.6) is in a linear form and the QL score function has a unique root under the given dependence structure, convergence of the computational algorithm can be seen in Corollary 5. More discussions about convergence can be found in Section 6 and the Supplementary Material.

3.2. Multiple jump sets identification

We now develop a sequential method to identify multiple jump sets based on the identification procedure developed in Section 3.1. Assume that by Algorithm

1, we have chosen k status vectors, say $\hat{\boldsymbol{\delta}}_1, \ldots, \hat{\boldsymbol{\delta}}_k$, in the first k stages, $k = 1, 2, \ldots$ Let $\boldsymbol{\Omega}^{(k+1)}$ denote a collection of candidate status vectors that has been updated at the $(k+1)$th-stage, and let $\boldsymbol{\psi}_\alpha^{(k+1)} \in \boldsymbol{\Omega}^{(k+1)}$ denote a given candidate status vector. To mitigate collinearity in the multiple jump-set model, we require $\hat{\boldsymbol{\delta}}_1, \ldots, \hat{\boldsymbol{\delta}}_k$ and $\boldsymbol{\psi}_\alpha^{(k+1)}$ to be disjoint. That is, each element in $\hat{\boldsymbol{\delta}}_1 + \cdots + \hat{\boldsymbol{\delta}}_k + \boldsymbol{\psi}_\alpha^{(k+1)}$ is less than or equal to one. How to partition $\hat{\boldsymbol{\delta}}_1, \ldots, \hat{\boldsymbol{\delta}}_k$ and $\boldsymbol{\psi}_\alpha^{(k+1)}$ such that they are disjoint can be seen in Section 4.

A multiple jump-set model for the mean response in terms of $\boldsymbol{\psi}_\alpha^{(k+1)}$ associated with $\hat{\boldsymbol{\delta}}_1, \ldots, \hat{\boldsymbol{\delta}}_k$ is given by

$$\boldsymbol{\theta}_{\boldsymbol{\psi}_\alpha^{(k+1)}; \hat{\boldsymbol{\delta}}_1, \ldots, \hat{\boldsymbol{\delta}}_k} = g^{-1}\left\{ \beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \xi_{\hat{\boldsymbol{\delta}}_j} \hat{\boldsymbol{\delta}}_j + \xi_{\boldsymbol{\psi}_\alpha^{(k+1)}} \boldsymbol{\psi}_\alpha^{(k+1)} \right\}, \tag{3.2}$$

where $\boldsymbol{\beta}$, $\xi_{\hat{\boldsymbol{\delta}}_1}, \ldots, \xi_{\hat{\boldsymbol{\delta}}_k}$, and $\xi_{\boldsymbol{\psi}_\alpha^{(k+1)}}$ are the coefficients for the covariates, the estimated status vectors $\hat{\boldsymbol{\delta}}_1, \ldots, \hat{\boldsymbol{\delta}}_k$, and the candidate status vector $\boldsymbol{\psi}_\alpha^{(k+1)}$, respectively. Let $\hat{\boldsymbol{\delta}}_j^c = 1 - \hat{\boldsymbol{\delta}}_j$, $j = 1, \ldots, k$, and $\bar{\boldsymbol{\psi}}_\alpha^{(k+1)} = 1 - \boldsymbol{\psi}_\alpha^{(k+1)}$. The complementary model then becomes

$$\boldsymbol{\theta}_{\bar{\boldsymbol{\psi}}_\alpha^{(k+1)}; \hat{\boldsymbol{\delta}}_1^c, \ldots, \hat{\boldsymbol{\delta}}_k^c} = g^{-1}\left\{ \beta_0^c + \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \xi_{\hat{\boldsymbol{\delta}}_j^c} \hat{\boldsymbol{\delta}}_j^c + \xi_{\bar{\boldsymbol{\psi}}_\alpha^{(k+1)}} \bar{\boldsymbol{\psi}}_\alpha^{(k+1)} \right\}, \tag{3.3}$$

where $\beta_0^c = \beta_0 + \sum_{j=1}^{k} \xi_{\hat{\boldsymbol{\delta}}_j} + \xi_{\boldsymbol{\psi}_\alpha^{(k+1)}}$. Note that models (3.2) and (3.3) are analogous to the generalized complementary models (2.4) and thus, the sequential identification procedure is the same as that for finding a single jump set.

4. Simulation Study

In this section, we conduct a simulation study to study consistency and convergence of the proposed method for irregularly gridded data. Additional simulations for regularly gridded data are presented in the Supplementary Material with related discussions given in Section 6. In particular, we introduce a partition method that can create disjoint (candidate) jump sets for models (3.2) and (3.3) associated with irregularly gridded data.

To simulate irregularly gridded data, we use the geographic structure of the poverty case study with $n = 535$ counties and five socio-economic explanatory variables, x_1 through x_5. Three collections of geographic cells, Δ_1, Δ_2, and Δ_3, are chosen to be the true jump sets with $|\Delta_1| = 20$, $|\Delta_2| = 50$, and $|\Delta_3| = 50$ (Fig. A2 of the Supplementary Material). Let $\delta_{k,i} = I[\boldsymbol{s}_i \in \Delta_k]$ denote the status variables associated with Δ_k, $i = 1, \ldots, 535$, $k = 1, 2, 3$. For the lth simulation run, we generate $\theta_{i,l}^* = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \xi_1 \delta_{1,i} + \xi_2 \delta_{2,i} + \xi_3 \delta_{3,i} + \epsilon_{i,l}$, where $\epsilon_{i,l}$ are simulated from a Gaussian distribution with mean zero, variance 0.0025, and correlation $\text{corr}(\epsilon_{i,l}, \epsilon_{j,l}) = 0.5 \exp(-0.005 \|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$. Also,

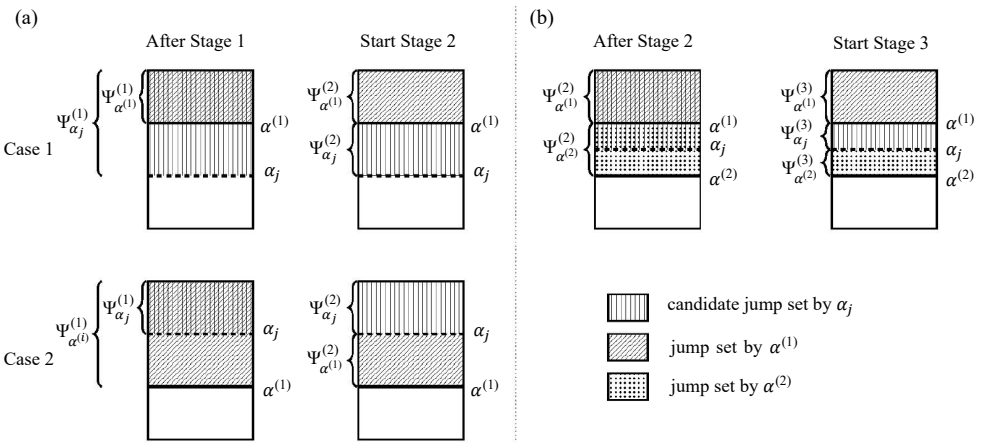

Figure 1. Examples for making disjoint jump sets by the partition method in various stages. Chosen thresholds $\alpha^{(1)}$ and $\alpha^{(2)}$ and candidate threshold α_j are used to segment observations. (a) The partition process associated with $\alpha^{(1)}$ chosen from Stage 1. (b) The partition process associated with $\alpha^{(1)}$ and $\alpha^{(2)}$ chosen from Stages 1 and 2, respectively.

we set the true regression coefficients to $\beta_0 = -1$, $\beta_1 = 1$, $\beta_2 = -0.5$, $\beta_3 = 1$, $\beta_4 = -5$, $\beta_5 = 1.5$, and the jump coefficients $\xi_1 = 3$, $\xi_2 = 2$, and $\xi_3 = 1$. The simulated poverty rates are then given by $Y^*_{i,l} = \exp(\theta^*_{i,l})/\{1 + \exp(\theta^*_{i,l})\}$, $i = 1, \ldots, 535$, and transformed to $Y^l_i = \log\{Y^*_{i,l}/(1 - Y^*_{i,l})\}$ (Sec. 5). There are a total of $L = 200$ simulation runs (or, replicates).

For jump set identification, we create an initial collection of candidate jump sets by $\Psi^{(1)}_{\alpha_j} = \{\boldsymbol{s}_i : \alpha_j \le Y_i\}$, $j = 1, \ldots, N$, where $\alpha_1 < \cdots < \alpha_N$ denote a collection of thresholds. Let $\boldsymbol{\psi}^{(1)}_{\alpha_j}$ denote a candidate status vector corresponding to the jump set $\Psi^{(1)}_{\alpha_j}$. Also, let $\mathcal{I}^{(1)} \equiv \{\alpha_1, \ldots, \alpha_N\}$ denote an initial collection of thresholds, and let $\boldsymbol{\Omega}^{(1)} \equiv \{\boldsymbol{\psi}^{(1)}_{\alpha_j} : j = 1, \ldots, N\}$ denote an initial collection of candidate status vectors. Algorithm 1 is then applied to identify a status vector from $\boldsymbol{\Omega}^{(1)}$ as an estimated status vector for the jump set. If there is no significant estimated status vector, then the identification procedure stops. Otherwise, a significant estimated status vector $\bar{\boldsymbol{\delta}}^{(1)}_1$ from $\boldsymbol{\Omega}^{(1)}$ is found and we continue as follows to ensure that the required condition for model (3.2) be satisfied.

Figure 1(a) illustrates how to update each status vector in the first stage so that they are disjoint in the second stage. First, with $\bar{\boldsymbol{\delta}}^{(1)}_1 \in \boldsymbol{\Omega}^{(1)}$, we have $\alpha^{(1)} \in \mathcal{I}^{(1)}$ such that $\bar{\boldsymbol{\delta}}^{(1)}_1 \equiv \boldsymbol{\psi}^{(1)}_{\alpha^{(1)}}$. Also, let $\alpha_j \in \mathcal{I}^{(2)}$ be a given threshold, where $\mathcal{I}^{(2)} = \mathcal{I}^{(1)} - \alpha^{(1)}$. Figure 1(a) depicts two possible relationships between $\alpha^{(1)}$ and α_j: $\alpha^{(1)} > \alpha_j$ (case 1) or $\alpha^{(1)} < \alpha_j$ (case 2), where $\boldsymbol{\Psi}^{(1)}_{\alpha^{(1)}}$ and $\boldsymbol{\Psi}^{(2)}_{\alpha^{(1)}}$ denote the corresponding estimated jump set associated with $\alpha^{(1)}$ in the first and second stages, respectively. Also, we let $\Psi^{(1)}_{\alpha_j}$ and $\Psi^{(2)}_{\alpha_j}$ denote the candidate jump set associated with α_j in the first and second stages, respectively. Thus, in case 1 ($\alpha^{(1)}$ larger), the estimated jump set associated with $\alpha^{(1)}$ is the same in both stages with $\Psi^{(1)}_{\alpha^{(1)}} \equiv \Psi^{(2)}_{\alpha^{(1)}} = \{\boldsymbol{s}_i : \alpha^{(1)} \le Y_i\}$, while the candidate jump set associated with α_j

is updated from $\Psi^{(1)}_{\alpha_j} = \{\boldsymbol{s}_i : \alpha_j \le Y_i\}$ to $\Psi^{(2)}_{\alpha_j} = \{\boldsymbol{s}_i : \alpha_j \le Y_i < \alpha^{(1)}\}$. In case 2 ($\alpha_j$ larger), the candidate jump set associated with α_j remains the same in both stages with $\Psi^{(1)}_{\alpha_j} \equiv \Psi^{(2)}_{\alpha_j} = \{\boldsymbol{s}_i : \alpha_j \le Y_i\}$, while the estimated jump set associated with $\alpha^{(1)}$ is updated from $\Psi^{(1)}_{\alpha^{(1)}} = \{\boldsymbol{s}_i : \alpha^{(1)} \le Y_i\}$ to $\Psi^{(2)}_{\alpha^{(1)}} = \{\boldsymbol{s}_i : \alpha^{(1)} \le Y_i < \alpha_j\}$.

Let $\boldsymbol{\psi}^{(2)}_{\alpha^{(1)}}$ and $\boldsymbol{\psi}^{(2)}_{\alpha_j}$ denote the corresponding status vectors for $\Psi^{(2)}_{\alpha^{(1)}}$ and $\Psi^{(2)}_{\alpha_j}$, respectively. Proceeding from the disjoint status vectors, we now build a multiple jump-set model associated with $\bar{\boldsymbol{\delta}}^{(2)}_1 \equiv \boldsymbol{\psi}^{(2)}_{\alpha^{(1)}}$ and $\boldsymbol{\psi}^{(2)}_{\alpha_j}$ for analyzing the simulated data. Let $\mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i}$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{535})'$. The multiple jump-set model associated with thresholds $\alpha^{(1)}$ and α_j is

$$\boldsymbol{\theta}_{\psi^{(2)}_{\alpha_j};\bar{\delta}^{(2)}_1} = g^{-1}\left\{\boldsymbol{\mu} + \xi_{\bar{\delta}^{(2)}_1}\bar{\boldsymbol{\delta}}^{(2)}_1 + \xi_{\psi^{(2)}_{\alpha_j}}\boldsymbol{\psi}^{(2)}_{\alpha_j}\right\}, \tag{4.1}$$

where $\xi_{\bar{\delta}^{(2)}_1}$ and $\xi_{\psi^{(2)}_{\alpha_j}}$ denote the jump coefficients associated with $\bar{\boldsymbol{\delta}}^{(2)}_1$ and $\boldsymbol{\psi}^{(2)}_{\alpha_j}$, respectively. Our identification procedure in Sections 3.1 and 3.2 for model (4.1) then searches over all thresholds $\alpha_j \in \mathcal{I}^{(2)}$. If there is a threshold, say $\alpha^{(2)}$, associated with another estimated status vector, then Figure 1(b) illustrates an example how to update the status vectors from the second to the third stage.

For the lth simulation, we create candidate thresholds $\alpha^l_j = \min\{Y^l_j : j = 1, \ldots, 535\} + 0.001(j-1)$, $j = 1, 2, \ldots$, such that α^l_j is less than $\max\{Y^l_j : j = 1, \ldots, 535\} + 0.001$. Our HLR method is applied to analyze the simulated data and is compared with an alternative approach by the HM-QL method (Lin and Zhu, 2020). If two estimated status vectors $\bar{\delta}_k$ and $\bar{\delta}_{k'}$ are close in the sense that the difference of the estimated coefficients $|\hat{\xi}_{\bar{\delta}_k} - \hat{\xi}_{\bar{\delta}_{k'}}|$ is less than 0.1, then they are considered to be the same. Also, if an estimated status vector has an estimated jump coefficient smaller than 0.1, then we leave out the estimated status vector. The final identified jump sets are sorted by their corresponding p-values, from the smallest to the largest. For the lth simulation, let $\bar{\Delta}^l_{i,\mathrm{LR}}$ and $\bar{\Delta}^l_{i,\mathrm{HM}}$ denote the corresponding identified jump sets with the ith smallest p-value by the HLR and HM-QL methods, respectively. $\bar{\Delta}^l_{i,\mathrm{LR}}$ and $\bar{\Delta}^l_{i,\mathrm{HM}}$ are used to be estimates for Δ_i, $i = 1, 2, 3$, by the HLR and HM-QL methods, respectively, at the lth simulation. For convenience, we also use $\bar{\Delta}_{i,\mathrm{LR}}$ and $\bar{\Delta}_{i,\mathrm{HM}}$ to denote $\bar{\Delta}^l_{i,\mathrm{LR}}$ and $\bar{\Delta}^l_{i,\mathrm{HM}}$, respectively. The average numbers of counties in Δ_i that have been classified into $\bar{\Delta}_{i,\cdot}$ are

$$T_{i,j} = L^{-1}\sum_{l=1}^{L} |\bar{\Delta}^l_{i,\cdot} \cap \Delta_j|, \ i, j = 1, 2, 3, \tag{4.2}$$

and thus, $T_{i,i}/|\Delta_i|$ represents a true positive rate of $\bar{\Delta}_{i,\cdot}$ for Δ_i (Tbl. 1). Further, we let $\bar{\Delta}_{4,\cdot}$ denote a collection of the identified jump sets other than $\bar{\Delta}_{1,\cdot}$, $\bar{\Delta}_{2,\cdot}$, and $\bar{\Delta}_{3,\cdot}$, and $\Delta_\emptyset$ denote a collection of counties without jump coefficients. That is, $\bar{\Delta}_{4,\cdot} = \cup_{i\ge 4}\bar{\Delta}_{i,\cdot}$ and $\Delta_\emptyset = D - \Delta_1 - \Delta_2 - \Delta_3$.

Table 1. The average numbers of counties in the true jump set Δ_i, $i = 1, 2, 3$, that were classified into $\bar{\Delta}_{j,\cdot}$ by the homogeneous likelihood ratio (HLR) and heterogeneity measure quasi-likelihood (HM-QL) methods. The simulation result is based on 200 replicates.

	HLR				HM-QL			
True	$\bar{\Delta}_{1,\mathrm{LR}}$	$\bar{\Delta}_{2,\mathrm{LR}}$	$\bar{\Delta}_{3,\mathrm{LR}}$	$\bar{\Delta}_{4,\mathrm{LR}}$	$\bar{\Delta}_{1,\mathrm{HM}}$	$\bar{\Delta}_{2,\mathrm{HM}}$	$\bar{\Delta}_{3,\mathrm{HM}}$	$\bar{\Delta}_{4,\mathrm{HM}}$
Δ_1	20.00	0.00	0.00	0.00	20.00	0.00	0.00	0.00
Δ_2	0.00	50.00	0.00	0.00	0.00	44.30	5.70	0.00
Δ_3	0.00	0.00	48.50	1.50	0.00	5.90	30.20	6.80
$\Delta_\emptyset$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2. Simulation results by homogeneous likelihood ratio (HLR) and heterogeneity measure quasi-likelihood (HM-QL) methods. Means and standard errors (in parentheses) for the estimated parameters with three jump sets $\Delta_1, \ldots, \Delta_3$, based on 200 replicates.

Variable	β_0	β_1	β_2	β_3	β_4	β_5	Δ_1	Δ_2	Δ_3
True	-1.0	1.0	-0.5	1.0	-5.0	1.5	3.0	2.0	1.0
HLR	-1.0	1.0	-0.5	1.0	-4.9	1.5	3.0	2.0	1.0
	(0.04)	(0.04)	(0.03)	(0.08)	(0.19)	(0.06)	(0.01)	(0.01)	(0.01)
HM-QL	-1.2	1.9	-0.6	1.5	-6.8	2.7	2.9	1.3	0.8
	(0.1)	(0.6)	(0.1)	(0.5)	(1.1)	(0.4)	(0.05)	(0.2)	(0.1)

Table 1 shows that our HLR method accurately identifies the jump sets Δ_1 and Δ_2 by $\bar{\Delta}_{1,\mathrm{LR}}$ and $\bar{\Delta}_{2,\mathrm{LR}}$, respectively, in all the simulation runs. Also, on average, the HLR method classifies 97% of Δ_3 into $\bar{\Delta}_{3,\mathrm{LR}}$, while a small proportion of Δ_3 are classified to additional estimated jump sets. On the other hand, although the HM-QL method can also accurately identify Δ_1 in all the simulation runs, it has a tendency to under-detect a jump set when the magnitude is relatively weak. For example, the HM-QL method misses about 14% of the hot spots in Δ_3, whose jump coefficient is equal to 1. We also test the difference of the true positive rates between the HLR and HM-QL methods by a traditional Z-test. Specifically, for $i = 1, 2, 3$, let $T_{i,i}^{\mathrm{LR}}$ and $T_{i,i}^{\mathrm{HM}}$ denote the corresponding values of $T_{i,i}$ by the HLR and HM-QL methods, respectively. And, let $P_i^{\mathrm{LR}} = T_{i,i}^{\mathrm{LR}}/|\Delta_i|$ and $P_i^{\mathrm{HM}} = T_{i,i}^{\mathrm{HM}}/|\Delta_i|$ denote the true positive rates for Δ_i by the HLR and HM-QL methods, respectively. Let $\mathrm{Diff}_i = P_i^{\mathrm{LR}} - P_i^{\mathrm{HM}}$ denote the difference of the true positive rates between the HLR and HM-QL methods for Δ_i, $i = 1, 2, 3$. From Table 1, Diff_2 and Diff_3 have Z-ratios of 2.5 (p-value$\doteq$ 0.01) and of 5.0 for (p-value$\doteq$ 0.00), respectively. This provides evidence that the HLR method outperforms the HM-QL method in identifying jump sets whose jump coefficients are moderately small, based on the Z-test.

Additionally, Table 2 provides the estimation result by averaging the estimated parameters across the $L = 200$ simulation runs for each method. From Table 2, the HLR method estimates the model parameters well with generally small biases and relatively stable standard errors. The HM-QL method, in

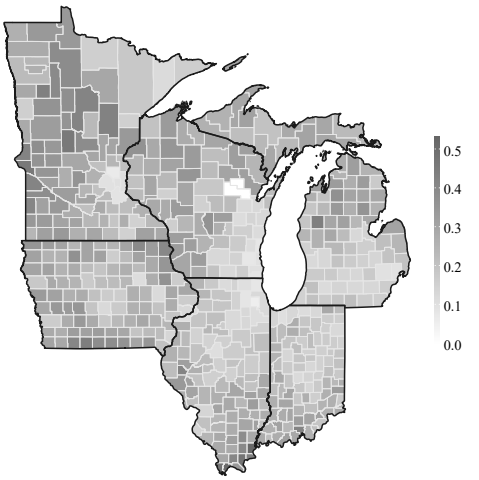

Figure 2. An image plot for the transformed poverty rates in the Midwest data.

contrast, tends to underestimate the jump coefficients for Δ_2 and Δ_3, leading to larger biases for the parameter estimates. One possible reason for this is that the HM-QL method uses a variable selection procedure to identify jump sets and thus, confounding effects could make serious impact when the number of covariates is large. This result agrees with the phenomenon that the HM-QL method has less power in identifying Δ_2 and Δ_3 than the HLR method.

5. Data Analysis for the Midwest Poverty

For illustration of the methodology developed in this paper, we examine poverty in relation to the social and economic factors in the US based on the 1960 census data in 535 counties of the five states in the Upper Midwest (Illinois, Indiana, Michigan, Minnesota, and Wisconsin). The response is a poverty rate, computed as the proportion of the county's population living below the poverty threshold. Let $\boldsymbol{s}_i$ denote the latitude and longitude of the centroid of county i, and let $Y_i^* \equiv Y^*(\boldsymbol{s}_i)$ denote the poverty rate. The observed poverty rate ranges are from 0.055 to 0.526 with a mean value of 0.245. As in Curtis et al. (2013), we take a logistic transformation of the poverty rate $Y_i = \log\{Y_i^*/(1 - Y_i^*)\}$.

The observed transformed poverty rates are mapped in Figure 2. A QQ-plot (not shown here) for the transformed data indicates that $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ follow a Gaussian distribution approximately. Thus in the data analysis, we adopt an identity link function. A quasi-deviance method (Lin, 2011) is applied and selects among the five explanatory variables, namely, the proportions of the county population employed in agriculture (x_1), manufacturing (x_2), services (x_3), finance, insurance, and real estate (FIRE) (x_4), and the proportion of African American (x_5). Let

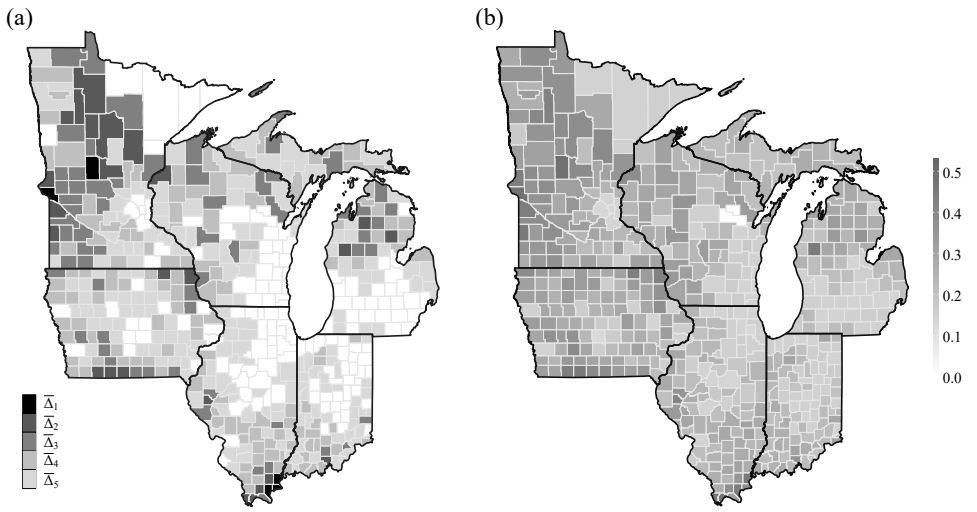

Figure 3. (a) A map for counties in the five identified jump sets, $\bar{\Delta}_1, \ldots, \bar{\Delta}_5$, from the data analysis. (b) An image plot for the estimated poverty rates from the final jump-set model.

$$\mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} \tag{5.1}$$

denote a marginal mean of Y_i associated with the explanatory variables. The estimates of the intercept β_0 and the slopes $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_5)'$ by a standard regression are given in Table A5 of Supplementary Material, and the slopes are all significant.

When setting five states as five known jump sets, we find that most estimates for the jump coefficients are not significant. This finding suggests that spatial pattern of jump sets would not follow the state boundaries, thereby nullifying the utility of standard spatial regimes. To identify more localized but hidden jump sets besides the explanatory variables, we apply our HLR method. First, we sort the transformed poverty rates, from the largest to smallest ones, to be $Y_{(1)} > Y_{(2)} > \cdots > Y_{(535)}$. Let $\alpha_j \equiv Y_{(j)}$ denote a candidate threshold. A candidate status variable associated with the threshold α_j for county i is given by $\psi_{\alpha_j,i} = I[\alpha_j \leq Y_i]$. Let $\boldsymbol{\psi}_{\alpha_j} = (\psi_{\alpha_j,1}, \ldots, \psi_{\alpha_j,535})'$ denote the status vector associated with α_j and let $\boldsymbol{\Omega} = \{\boldsymbol{\psi}_{\alpha_j} : j = 1, \ldots, 535\}$ denote the collection of the resulting status vectors. We then use the identification procedure similar to the one used in the simulation to sequentially identify the multiple jump sets by the HLR method. Additionally, similar to the simulation, we also combine two identified jump sets if difference between the corresponding estimated jump coefficients is not significant.

By the HLR method, five jump sets, namely, $\bar{\Delta}_1, \ldots, \bar{\Delta}_5$ are identified. Figure 3(a) shows locations of the five jump sets with $|\bar{\Delta}_1| = 5$, $|\bar{\Delta}_2| = 25$, $|\bar{\Delta}_3| = 63$, $|\bar{\Delta}_4| = 126$, and $|\bar{\Delta}_5| = 176$. Note that the five jump sets include about 74%

of total counties, and as can be seen from Figure 3(a), the identified counties are distributed quite randomly in the study area. These phenomena may thus render traditional clustering methods infeasible to find geographic clusters. For example, we also apply the spatial scan (*SatScan*) method to analyze the Midwest data, but *SatScan* does not yield any geographic cluster. Thus, both approaches appear to be under-powered. The final estimated model for the marginal mean response is thus given by

$$\hat{\theta}_i = \hat{\mu}_i + \sum_{j=1}^{5} \hat{\xi}_{\bar{\Delta}_j} \delta_{\bar{\Delta}_j}(\boldsymbol{s}_i), \tag{5.2}$$

where $\delta_{\bar{\Delta}_j}(\boldsymbol{s}_i)$ is the status variable for $\bar{\Delta}_j$, and $\hat{\xi}_{\bar{\Delta}_j}$ is an estimate for the jump coefficient associated with $\bar{\Delta}_j$. The parameter estimates are given in Table A5 of the Supplementary Material, along with the standard errors by Corollary 2 of the Appendix. Additionally, the estimated spatial correlation function is $\hat{\rho}_{i,j} = 0.62 \exp(-0.005\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$, with a moderate spatial correlation ranging from 0.48 to 0.62 for counties within 50 km of each other. The estimated poverty rates $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_{535})'$ by the HLR method are mapped in Figure 3(b), which are similar to the observed poverty rates in Figure 2.

To compare the models with and without jump sets (5.1)–(5.2), we compute a weighted least squares (WLS) $(\boldsymbol{Y} - \hat{\boldsymbol{\theta}})'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{V}}$ is an estimate of the covariance matrix for $\boldsymbol{Y}$. As can be seen from Table A5 of the Supplementary Material, accounting for jump sets in the model substantially improves the model fit based on the WLS values. The regression coefficients and the jump coefficients for the five jump sets are all significant. Five jump sets are identified and the counties in the same jump set tend to be close to one another geographically. The counties in the jump sets $\bar{\Delta}_1$ and $\bar{\Delta}_2$ have the higher poverty rates and tend to concentrate in the northern and southern parts of the study region. Many counties along the shores of the Great Lakes are in the complement of the jump sets with low poverty rates, which is plausible due to the positive association with strong, stable manufacturing jobs in the area during this period of pre-deindustrialization. The spatial pattern of jump sets suggests smaller scale, more localized forces are at play in generating county poverty rates. Ultimately, results direct future research to examine shared attributes and processes among the five collections of place-types, perhaps including the ways in which industrial and racial forces interact.

Further, for the study region as a whole, the proportions of agriculture and service employment as well as the proportion of African Americans are positively related to poverty, whereas those of manufacturing and FIRE have a negative relation. This pattern reflects a greater vulnerability to poverty among places that were more reliant upon agriculture and services as compared to places more reliant up on manufacturing and the FIRE industries. Above

and beyond the influence of industry, poverty rates were higher in counties with larger concentrations of African Americans (x_5), a racialized group historically marginalized in American society.

6. Conclusions and Discussion

Here we have developed novel methodology for simultaneous regression and jump-set analysis for the case study of poverty history in the Upper Midwest of the US. Since the hidden jump sets are unknown, the proposed generalized complementary models are different from the traditional generalized linear models. We have also proposed a partition approach associated with the complementary models to create candidate jump sets and developed a novel homogeneity measure from an approximation of the log-QL likelihood ratio to connect jump set identification and regression analysis. Since the homogeneity measure is in a quadratic form and the proposed score function has a unique root, the three-step computational algorithm achieves convergence fairly quickly. Particularly, in the simulation study shown in the Supplementary Material, the HLR method for Gaussian responses takes only one iteration toward convergence almost every time (Tbl. A1). A QL Z-ratio test shown in Section 3.1 is also proposed to evaluate whether a hidden jump set exists.

In addition, we have established a large-sample property (consistency) for the HLR method in a series of theorems. A simulation study in the Supplementary Material shows that the HLR method is also consistent in jump set identification rates and parameter estimation for Gaussian and Poisson responses (Tbls. A1 and A2). That is, as sample size increases, the identification accuracy for jump sets increases and the sample variance for estimates decreases. The finite-sample properties in the simulation study have supported the theory. Thus, the proposed method based on the QL homogeneity measure is both theoretically justified and computationally feasible. In another simulation study (not shown here), we have found that the more similar candidate jump sets are to the true jump set, the closer the corresponding homogeneity measures are to zero. For example, for the jump set $|\Delta_3|$ shown in Figure A2 of the Supplementary Material with $|\Delta_3| = 150$, an average value of the simulated homogeneity measures for Δ_3 over 200 replicates is about 1.1, while those for candidate jump sets with numbers of counties to be 120, 130, and 140 are 68, 42, and 20, respectively.

For the HLR method, the computation time would depend on the number of candidate jump sets. For example, in analysis of the Midwest data, 535 candidate jump sets were considered, and the HLR method would take 4 hours to get convergence. However, in the simulation for 12×12 grids of the Gaussian responses, it would take only 20 minutes for one simulation setting (500 replicates). Additionally, we have shown that estimates by the HLR method have asymptotic normality, and used a WLS error as a criterion for model selection in

the data analysis. In a simulation study shown in the Supplementary Material that has a high proportion of counties included in the jump sets, we have found that the HLR method can identify all the true jump sets accurately and give unbiased estimates for most coefficients (Tbls. A3 and A4). To further evaluate whether the WLS error is suitable for model selection, we have computed the WLS error in the simulation. The simulation results show that an average of WLS errors for the (final) jump set model is about 0.5, while that for the traditional regression model is about 50 (Tbl. A4). These results support the use of WLS errors for the data analysis. To further improve the accuracy and efficiency of our methodology, we could consider more flexible ways of creating candidate jump sets. Another natural extension of our methodology would be simultaneous identification of jump sets in space and change points in time by creating spatio-temporal status vectors, which we leave for future investigation.

Supplementary Material

Technical details, extra simulation studies and computer codes can be found in the Supplementary Material.

Acknowledgments

The authors would like to thank editor, associate editor, and two anonymous referees for their constructive comments toward improving the paper. This material is based upon work supported by and conducted during service at the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix: Assumptions and Theorems

To ensure a unique root for a score function of the QL function, we make some assumptions for the QL function to exist.

Assumption 1. *(a) The mean functions satisfy* $\max\{|\theta_i|^4 : i = 1, \ldots, n\}$ *is finite. (b) The explanatory variables* $\boldsymbol{x}_i$ *are not multiples of a binary variable. (c) The link function* $g(\cdot)$ *is one-to-one. (d) The first- and second-order derivatives of the link function* $g(\cdot)$ *are continuous. (e) There exists a smooth function* $V(\cdot)$ *such that* $var(Y_i) = V(\theta_i, \sigma)$*, where* σ *is a nuisance parameter.*

We next impose mixing conditions for the responses to ensure the validity of the QL estimation for the jump-set model. Let $\Xi \subseteq D$ and let $\boldsymbol{Y}_\Xi = \prod_{\boldsymbol{s}_i \in \Xi} Y(\boldsymbol{s}_i)$. Let $\rho_{k,l}(h) = \sup\big\{\, |\mathrm{corr}(\boldsymbol{Y}_{\Xi_1}, \boldsymbol{Y}_{\Xi_2})| : |\Xi_1| \le k, |\Xi_2| \le l, d(\Xi_1, \Xi_2) \ge h \big\}$, where $\mathrm{corr}(\cdot, \cdot)$ denotes a correlation function, and $d(\Xi_1, \Xi_2) = \inf\{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| : \boldsymbol{s}_i \in \Xi_i\}$ (Lin, 2008).

Assumption 2. *The mixing coefficient $\rho_{k,l}(h)$ satisfies the following conditions: (a) $\rho_{1,1}(h) = O(h^{-2-k_1})$ for some $k_1 > 0$. (b) $\rho_{k,l}(h) = o(h^{-2})$ for $k + l \leq 4$. (c) $\rho_{1,1}(h)$ is positive definite.*

By Assumption 2(a), we have $\sum_{i=1}^{n} \Lambda_{i,j} = O(1)$ for all $j = 1, \ldots, n$ and thus it is reasonable to make the following assumption.

Assumption 3. *The information matrix $-n^{-1}(\nabla_{\boldsymbol{\lambda}_\delta}\boldsymbol{\theta}_\delta)'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}_\delta}(\nabla_{\boldsymbol{\lambda}_\delta}\boldsymbol{\theta}_\delta)$ converges to a positive-definite matrix $\boldsymbol{I}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$ as $n \to \infty$.*

Let $\xrightarrow{P}$ and $\xrightarrow{\mathcal{L}}$ denote convergence in probability and in distribution, respectively, as $n \to \infty$. Also, let $\boldsymbol{o}^*(n^q)$ denote a vector containing either 0 or 1 with a sum on the order $o(n^q)$ as $n \to \infty$. Let $\boldsymbol{\psi}_0$ be a given status vector with $\boldsymbol{\psi}_0 = \boldsymbol{\delta} + \boldsymbol{o}^*(n^{1/2})$. We have the following results for asymptotic properties of the HLR method.

Theorem 1. *Suppose $n_\delta = O(n)$ and Assumptions 1–3 hold. Then, at each iteration $m = 0, 1, 2, \ldots$ of Algorithm 1, the QL estimates $\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}$ are consistent in the sense that $\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} \xrightarrow{P} \boldsymbol{\lambda}_\delta$, as $n \to \infty$.*

Theorem 1 states that, when a candidate status vector and the true status vector are asymptotically equivalent, the QL estimates associated with the corresponding candidate model are consistent. In more typical iterative estimation methods, the consistency of regression parameter estimates hinges on the consistency of covariance parameter estimates (Guyon, 1995). Here, in contrast, the estimates of the regression coefficients and the jump coefficient can be consistent, even if the covariance parameter estimates are biased.

Corollary 1. *Under the assumptions of Theorem 1, we have $\hat{\boldsymbol{\lambda}}_{\psi}^{(m)} \xrightarrow{P} \boldsymbol{\lambda}_\delta$ and $\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)} \xrightarrow{P} \boldsymbol{\lambda}_{\delta^c}$ if and only if $\boldsymbol{\psi} \equiv \boldsymbol{\psi}_0 = \boldsymbol{\delta} + \boldsymbol{o}^*(n^{1/2})$, where $\boldsymbol{\lambda}_{\delta^c} = (\beta_0 + \xi_\delta, \boldsymbol{\beta}', -\xi_\delta)'$.*

The asymptotic normality of $\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}$ and $\hat{\boldsymbol{\lambda}}_{\psi_0^c}^{(m)}$ can be established on the consistency in the following Corollary 2.

Corollary 2. *Under the assumptions of Theorem 1, we have, as $n \to \infty$, $n^{1/2}\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} \xrightarrow{\mathcal{L}} N\left(\boldsymbol{\lambda}_\delta, \boldsymbol{I}^{-1}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}^\dagger)\right)$ and $n^{1/2}\hat{\boldsymbol{\lambda}}_{\psi_0^c}^{(m)} \xrightarrow{\mathcal{L}} N\left(\boldsymbol{\lambda}_{\delta^c}, \boldsymbol{I}^{-1}(\boldsymbol{\lambda}_{\delta^c}; \boldsymbol{\tau}^\dagger)\right)$ where $\boldsymbol{I}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}^\dagger)$ is given in Assumption 3. Moreover, if the selected variogram model is correctly specified, then, as $n \to \infty$, $n^{1/2}\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} \xrightarrow{\mathcal{L}} N\left(\boldsymbol{\lambda}_\delta, \boldsymbol{I}^{-1}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})\right)$, and $n^{1/2}\hat{\boldsymbol{\lambda}}_{\psi_0^c}^{(m)} \xrightarrow{\mathcal{L}} N\left(\boldsymbol{\lambda}_{\delta^c}, \boldsymbol{I}^{-1}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})\right)$.*

Let $n_\psi = \sum_{i=1}^{n} \psi_i$ and $n_\delta = \sum_{i=1}^{n} \delta_i$. The following Theorems 2 and 3 show that the QL homogeneity measure (2.6) has a (unique) minimum value when the candidate jump set is (asymptotically) equal to the true jump set. These results consequently ensure the selection consistency for the proposed jump-set identification procedure.

Theorem 2. *Suppose* $n_\delta = O(n)$, $n_\psi = O(n)$, *and Assumptions* 1–3 *hold. Then, at each iteration* $m = 0, 1, \ldots$ *of Algorithm* 1, *we have an inequality* $\min_{\boldsymbol{\psi}\in\boldsymbol{\Omega}} H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)})\} \geq 0$, *with probability one. In addition, the equality (asymptotically) holds if and only if* $\boldsymbol{\psi} \equiv \boldsymbol{\psi}_0$; *that is,* $H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)})\} = o_p(n^{-1})$ *if and only if* $\boldsymbol{\psi} = \boldsymbol{\delta} + \boldsymbol{o}^*(n^{1/2})$.

Assume that $\boldsymbol{\Omega}$ (asymptotically) contains the true status vector $\boldsymbol{\delta}$ in the sense that at least one status vector $\boldsymbol{\psi}_0$ is in $\boldsymbol{\Omega}$. By Theorem 2, $H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi_0^c}^{(m)})\}$ is (asymptotically) equal to zero, where $\boldsymbol{\psi}_0^c = 1 - \boldsymbol{\psi}_0$. This implies that $\boldsymbol{\psi}_0$ minimizes $H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)})\}$, which gives existence of $\min_{\boldsymbol{\psi}\in\boldsymbol{\Omega}} H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)})\}$. Recall that the estimated status vector at the mth iteration of Algorithm 1 is denoted by $\hat{\boldsymbol{\delta}}^{(m)} = \min_{\boldsymbol{\psi}\in\boldsymbol{\Omega}} H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)})\}$. Then, we have the following result about the consistency of the jump-set selection.

Theorem 3. *Suppose that* $\boldsymbol{\Omega}$ *(asymptotically) contains the true status vector* $\boldsymbol{\delta}$ *and that the assumptions of Theorem* 2 *hold. We have, at each iteration* $m = 0, 1, \ldots$ *of Algorithm* 1, $\hat{\boldsymbol{\delta}}^{(m)} = \min_{\boldsymbol{\psi}\in\boldsymbol{\Omega}} H\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)})\}$ *if and only if* $\hat{\boldsymbol{\delta}}^{(m)} = \boldsymbol{\delta} + \boldsymbol{o}^*(n^{1/2})$.

Theorem 3 states that the estimated status vector from the minimizer of the homogeneity measure is asymptotically equivalent to the true status vector. That is, Algorithm 1 asymptotically selects the true jump set and thus is consistent in the jump-set identification. We consider the asymptotic behavior of the case that there is no jump set (i.e., $\boldsymbol{\delta} = \mathbf{0}$).

Theorem 4. *Suppose that there is no jump set and Assumptions* 1–3 *hold. Then, the following results hold.* (a) The difference between the two jump coefficient estimates is asymptotically equivalent such that $(\hat{\xi}_{\psi} + \hat{\xi}_{\psi^c}) \xrightarrow{P} 0$ for any $\boldsymbol{\psi} \in \boldsymbol{\Omega}$, as $n \to \infty$. (b) The test statistic Z_ψ follows the standard normal distribution.

By Theorem 4, when there is no jump set, the test statistic Z_ψ is asymptotically normal and thus an approximate normal test can be performed.

Finally, we provide theorems for convergency of the proposed methods.

Theorem 5. *Under Assumptions* 1–3, *the derivative of* $n^{-1}\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$ *converges in probability to the positive-definite matrix* $\boldsymbol{I}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$ *of Assumption 3 as* $n \to \infty$. *Furthermore, the derivative of the QL function is uniformly bounded in probability as* $n \to \infty$. *That is,* $\|\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})\| \leq M$ *in probability as* $n \to \infty$.

Corollary 3. *Under Assumptions* 1–3, $\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$ *satisfies the Lipschitz condition in probability as* $n \to \infty$. *That is,* $\|\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}) - \dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta^*; \boldsymbol{\tau})\| \leq l\|\boldsymbol{\lambda}_\delta - \boldsymbol{\lambda}_\delta^*\|$ *in probability for some* $l > 0$ *as* $n \to \infty$.

Corollary 4. *Under Assumptions* 1–3, *a Newton iteration of the QL function is globally convergent in probability as* $n \to \infty$.

Corollary 5. *Under Assumptions* 1–3*, the iterative procedure of Algorithm* 1 *converges in probability as* $n \to \infty$.

The proofs are given in the Supplementary Material.

References

Argyros, I. K. (2008). *Convergence and Applications of Newtwon-type Iterations*. Springer, New York.

Curtis, K.J., Reyes, P.E., O'Connell, H. and Zhu, J. (2013). Assessing the spatial concentration and temporal persistence of poverty: Industrial structure, racial/ethnic composition, and the complex links to poverty. *Spatial Demography* **1**, 178–194.

Cressie, N. (1993). *Statistics for Spatial Data*. Revised Edition. Wiley, New York.

Guyon, X. (1995). *Random Fields on a Network*. Springer, New York.

Karimireddy, S. P., Stich, S. U. and Jaggi, M. (2019). Global linear convergence of Newton's method without strong-convexity or Lipschitz gradients. *arXiv:1806.00413*.

Li, B. (1993). A deviance function for the quasi-likelihood method. *Biometrika* **80**, 741–753.

Lin, P.-S. (2008). Estimating equations for spatially correlated data in multi-dimensional space. *Biometrika* **95**, 847–858.

Lin, P.-S. (2011). Quasi-deviance functions for spatially correlated data. *Statistica Sinica* **21**, 1785–1806.

Lin, P.-S. and Zhu, J. (2020). A heterogeneity measure for cluster identification with application to disease mapping. *Biometrics* **76**, 403–413.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. 2nd Edition. Chapman & Hall, Cambridge.

Zhang, H. (1998). Classification trees for multiple binary responses. *Journal of American Statisticians Society* **93**, 180–193.

(Received June 2022; accepted July 2023)