Supplementary Material for

A Homogeneity Likelihood Ratio Measure

for Hidden Jump-sets in Generalized Spatial Regression

Pei-Sheng Lin Jun Zhu Katherine J. Curtis

1 Technical Details

1.1 Assumptions

We first briefly introduce assumptions and theorems that are developed for the proposed method.

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.*

Assumption 1(b) ensures identifiability between the regression coefficients and status vectors, while the remainder of Assumption 1 are common regularity conditions for generalized linear models. We next impose some 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 $\rho_{k,l}(h) = \sup\big\{ |\text{corr}(\boldsymbol{Y}_{\Xi_1}, \boldsymbol{Y}_{\Xi_2})| : |\Xi_1| \leq k, |\Xi_2| \leq l, d(\Xi_1, \Xi_2) \geq h \big\}$, where $\text{corr}(\cdot, \cdot)$ denotes a correlation function, and $d(\Xi_1, \Xi_2) = \inf\{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| : \boldsymbol{s}_i \in \Xi_i\}$ (see, e.g., 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.*

Assumption 2 holds for the case of spatial independence and various spatial correlation models, including the exponential and spherical models. By Assumption 2(a), we have $\sum_{i=1}^{n} \Lambda_{i,j} = O(1)$

for all $j = 1, \ldots, n$ and thus, $(\nabla_{\boldsymbol{\lambda}_\delta} \boldsymbol{\theta}_\delta)' \boldsymbol{\Lambda}_{\boldsymbol{\lambda}_\delta} (\nabla_{\boldsymbol{\lambda}_\delta} \boldsymbol{\theta}_\delta) = \boldsymbol{O}(n)$. It is therefore 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* $n \to \infty$.

1.2 Theorems

To investigate the asymptotic properties, we consider an increasing domain. 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 $B(\boldsymbol{s}_i, r)$ denote a ball with center $\boldsymbol{s}_i$ and radius r. Under suitable regularity conditions, we first establish the convergence of $\hat{\boldsymbol{\tau}}$ from (2.9).

Proposition 1 *Suppose* $|B(\boldsymbol{s}_i, r) \cap D| = O(r^2)$ *for* $i = 1, \ldots, n$ *and Assumptions 1–2 hold. Then,* $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^\dagger + \boldsymbol{O}_p(n^{-1})$ *for some constants* $\boldsymbol{\tau}^\dagger$*. Furthermore, if the selected variogram model* $\gamma_{\boldsymbol{\tau}}(h)$ *is correctly specified, then* $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau} + \boldsymbol{O}_p(n^{-1})$*, where* $\boldsymbol{\tau}$ *is the true vector of covariance parameters.*

Let $n_\psi = \sum_{i=1}^n \psi_i$ and $n_\delta = \sum_{i=1}^n \delta_i$. Also, 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. Here, in contrast, the estimates of the regression coefficients and the jump coefficient can be consistent, even if the covariance parameter estimates are biased.

Next, we study the asymptotic behavior of the jump coefficient estimate $\hat{\xi}_\psi$ for an arbitrary status vector $\boldsymbol{\psi}$. Let $\Delta_\psi = \{j : \psi_j = 1\}$ and $\Delta_\psi^c = \{j : \psi_j = 0\}$ denote the jump set and the complement set of Δ_ψ, respectively, associated with $\boldsymbol{\psi}$. At the mth iteration of Algorithm 1, let

$\hat{\boldsymbol{\lambda}}_\psi^{(m)} = (\hat{\beta}_0^{(m)}, \hat{\boldsymbol{\beta}}_\psi^{(m)}, \hat{\xi}_\psi^{(m)})'$ denote the QL estimate for $\boldsymbol{\lambda}_\psi$. Let $\hat{\boldsymbol{\theta}}_\psi^{(m)} = g^{-1}\{\hat{\beta}_0^{(m)} + \boldsymbol{X}\boldsymbol{\beta}_\psi^{(m)} + \hat{\xi}_\psi^{(m)}\boldsymbol{\psi}\}$ with the jth element $\hat{\theta}_{j;\psi}^{(m)}$. Let $\hat{\boldsymbol{\theta}}_{\psi^c}^{(m)} = g^{-1}\{(\hat{\beta}_0^{(m)} + \hat{\xi}_\psi^{(m)}) + \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\psi^c}^{(m)} - \hat{\xi}_{\psi^c}^{(m)}\boldsymbol{\psi}^c\}$ with the jth element $\hat{\theta}_{j;\psi^c}^{(m)}$. Also, let $\boldsymbol{V}\{\hat{\boldsymbol{\lambda}}_\psi^{(m)}; \hat{\boldsymbol{\tau}}^{(m)}\}$ denote the working covariance matrix evaluated at the QL estimates and let $\boldsymbol{\Lambda}\{\hat{\boldsymbol{\lambda}}_\psi^{(m)}; \hat{\boldsymbol{\tau}}^{(m)}\}$ denote the inverse of $\boldsymbol{V}\{\hat{\boldsymbol{\lambda}}_\psi^{(m)}; \hat{\boldsymbol{\tau}}^{(m)}\}$ with the (i,j)th element denoted as $\hat{\Lambda}_{i,j;\psi}^{(m)}$.

Proposition 2 *Let $\boldsymbol{\psi}$ be an arbitrary status vector satisfying $n_\psi = O(n)$. Under Assumptions 1–3, we have, as $n \to \infty$,*

$$\sum_{j\in\Delta_\psi}\sum_{i=1}^{n} \hat{\theta}_{j;\psi}^{(m)}\hat{\Lambda}_{j,i;\psi}^{(m)}\{\beta_0 + \xi_\delta\delta_i - \hat{\beta}_0^{(m)} - \hat{\xi}_\psi^{(m)}\psi_i\} = O_p(n^{1/2}),$$

$$\text{and} \quad \sum_{j\in\Delta_\psi^c}\sum_{i=1}^{n} \hat{\theta}_{j;\psi^c}^{(m)}\hat{\Lambda}_{j,i;\psi^c}^{(m)}\{(\beta_0 + \xi_\delta) - \xi_\delta\delta_i^c - (\hat{\beta}_0^{(m)} + \hat{\xi}_\psi^{(m)}) + \hat{\xi}_{\psi^c}^{(m)}\psi_i^c\} = O_p(n^{1/2}).$$

There are several remarks for Proposition 2. First, unlike Theorem 1, no assumption about the size of the true jump set is made and thus, Proposition 2 can be applied to the case of no jump set. Second, when a candidate status vector $\boldsymbol{\psi}$ is mis-specified, Proposition 2 states that $\hat{\xi}_\psi^{(m)}$ is biased. This result reveals that, if $\hat{\xi}_\psi^{(m)}$ is consistent, then $\boldsymbol{\psi}$ is asymptotically equivalent to $\boldsymbol{\delta}$ (on the order of $n^{1/2}$). By Theorem 1 and Proposition 2, the following Corollary 1 holds.

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 Proposition 1 and 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 $n^{1/2}\hat{\boldsymbol{\lambda}}_{\psi_0^c}^{(m)} \xrightarrow{\mathcal{L}} N\left(\boldsymbol{\lambda}_{\delta^c}, \boldsymbol{I}^{-1}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})\right)$.*

Further, the following Theorems 2 and 3 show that the QL homogeneity measure (2.5) 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* $\min_{\boldsymbol{\psi} \in \boldsymbol{\Omega}} H \left\{ \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)}) \right\} \geq 0$, *with probability one. In addition, the equality (asymptotically) holds if and only if* $\boldsymbol{\psi} \equiv \boldsymbol{\psi}_0$*; that is,* $H \left\{ \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)}) \right\} = o_p(n^{-1})$ *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, $H \left\{ \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi_0^c}^{(m)}) \right\}$ is (asymptotically) equal to zero, where $\boldsymbol{\psi}_0^c = 1 - \boldsymbol{\psi}_0$. This implies that $\boldsymbol{\psi}_0$ minimizes $H \left\{ \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)}) \right\}$, which gives existence of $\min_{\boldsymbol{\psi} \in \boldsymbol{\Omega}} H \left\{ \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)}) \right\}$.

Also, recall that the estimated status vector at the mth iteration of Algorithm 1 is denoted by $\hat{\boldsymbol{\delta}}^{(m)} = \min_{\boldsymbol{\psi} \in \boldsymbol{\Omega}} H \left\{ \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)}) \right\}$. 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,* $\hat{\boldsymbol{\delta}}^{(m)} = \min_{\boldsymbol{\psi} \in \boldsymbol{\Omega}} H \left\{ \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)}) \right\}$ *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. Finally, 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. The proofs of Propositions 1–2, Theorems 1–4, and Corollaries 1–2 above are given in the next section.

Let $\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}) = \partial \boldsymbol{Q}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}) / \partial \boldsymbol{\lambda}_\delta$ denote a derivative matrix of $\boldsymbol{Q}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$ with respect to $\boldsymbol{\lambda}_\delta$. Below we show convergence for the proposed estimation method.

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,* $\|\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau})\|\le M$ *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,* $\|\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau})-\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}^*_\delta;\boldsymbol{\tau})\|\le l\|\boldsymbol{\lambda}_\delta-\boldsymbol{\lambda}^*_\delta\|$ *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* $n\to\infty$.

1.3 Proofs of Theorems

1.3.1 Derivation for the homogeneity measure

Let $\ell\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)\}$ and $\ell\{\boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c})\}$ denote the corresponding log-QL functions for $\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi);\boldsymbol{\tau}\}$ and $\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c});\boldsymbol{\tau}\}$, respectively. Since it is tenuous to have an explicit expression for the log-QL function when data are correlated, we apply a first-order Taylor series to expand $\Upsilon\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)\}$ at the reference point $\boldsymbol{\lambda}^+$. An approximation for the log-QL function can thus be given by

$$\ell\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)\}\approx(1/2)\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)-\boldsymbol{\theta}(\boldsymbol{\lambda}^+)\}'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}^+}[\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi)-\boldsymbol{\theta}(\boldsymbol{\lambda}^+)\}+2\{\boldsymbol{Y}-\boldsymbol{\theta}(\boldsymbol{\lambda}^+)\}], \tag{1.1}$$

Let $\hat{\boldsymbol{\lambda}}_\psi$ and $\hat{\boldsymbol{\lambda}}_{\psi^c}$ denote the QL estimates of $\boldsymbol{\lambda}_\psi$ and $\boldsymbol{\lambda}_{\psi^c}$, respectively. The log-QL ratio can thus be expressed as a difference between the log-QL functions under $\hat{\boldsymbol{\lambda}}_\psi$ and $\hat{\boldsymbol{\lambda}}_{\psi^c}$ by $H^0\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi),\boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c})\}=\ell\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi)\}-\ell\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}$. We can thus approximate $H^0\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi),\boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c})\}$ by

$$(1/2)\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi)-\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}^+}\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi)-\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}-\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi)-\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}^+}\{\boldsymbol{Y}-\boldsymbol{\theta}(\boldsymbol{\lambda}^+)\}.$$

We choose $\boldsymbol{\lambda}^+$ to satisfy the estimating equation $\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi)-\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\}'\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\lambda}^+}\{\boldsymbol{y}-\boldsymbol{\theta}(\boldsymbol{\lambda}^+)\}=0$, where $\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\lambda}^+}=(1/2)(\boldsymbol{\Lambda}_{\hat{\boldsymbol{\lambda}}_\psi}+\boldsymbol{\Lambda}_{\hat{\boldsymbol{\lambda}}_{\psi^c}})$, and thus $H^0\{\boldsymbol{\theta}(\boldsymbol{\lambda}_\psi),\boldsymbol{\theta}(\boldsymbol{\lambda}_{\psi^c})\}$ can be further simplified to the homogeneity measure. □

1.3.2 Proof of Proposition 1

We first show Proposition 1 holds for spatial data on a regular grid with size $m_1 \times m_2$. Recall that the number of pairs at distance h can be approximated by $|N(h)| \doteq 2m_1m_2h$ (Fingleton, 1983).

Let $\boldsymbol{\Gamma} = \text{var}(\hat{\boldsymbol{\gamma}})$ denote the covariance matrix of $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}(1), \ldots, \hat{\gamma}(M_n))'$ with the (i,j)th element $\Gamma_{i,j}$. Also, let h_n denote a sequence satisfying $\rho_{i,j}(h_n)(M_n)^{1/2} \to 0$ and $(h_n)^{-2}(M_n)^{1/2} \to \infty$, as $n \to \infty$. Note that $M_n = O(n)$. It can be shown that

$$\sum_i \sum_j \Gamma_{i,j} = \left[\sum_{i \le h_n} \sum_{j \le h_n} \text{cov}\{\hat{\gamma}(i), \hat{\gamma}(j)\} \right] \{1 + o(1)\}. \tag{A1}$$

Next, let $A_n = \sum_{i \le h_n} \sum_{j \le h_n} \text{cov}\{\hat{\gamma}(i), \hat{\gamma}(j)\}$. Then there exists some $C^* > 0$ such that

$$A_n \le C^* n^{-2} \sum\sum_{\|\boldsymbol{s}_k - \boldsymbol{s}_l\| \le h_n, \|\boldsymbol{s}_{k'} - \boldsymbol{s}_{l'}\| \le h_n} \text{corr}(\epsilon_k \epsilon_l, \epsilon_{k'} \epsilon_{l'}) \{1 + o(n^{-1/4})\}. \tag{A2}$$

In (A2), the correlation $\text{corr}(\epsilon_k \epsilon_l, \epsilon_{k'} \epsilon_{l'})$ can be separated into two parts: (i) if $\|\boldsymbol{s}_k - \boldsymbol{s}_{k'}\| = h \ge 3h_n$, then $\text{corr}(\epsilon_k \epsilon_l, \epsilon_{k'} \epsilon_{l'}) \le \alpha_{2,2}(h - 2h_n)$, and (ii) if $\min\{\|\boldsymbol{s}_k - \boldsymbol{s}_{k'}\|, \|\boldsymbol{s}_k - \boldsymbol{s}_l\|, \|\boldsymbol{s}_l - \boldsymbol{s}_{l'}\|\} = h \le 3h_n$, then $\text{corr}(\epsilon_k \epsilon_l, \epsilon_{k'} \epsilon_{l'}) \le k_0 \alpha_{1,3}(h)$ for some $k_0 > 0$. Thus, (A2) becomes

$$A_n \le C n^{-2} (h_n)^4 \left\{ \sum_{h=3h_n}^{\infty} \alpha_{2,2}(h - 2h_n) + \sum_{h=0}^{3h_n} \alpha_{1,3}(h) \right\}, \tag{A3}$$

for some $C > 0$, which is of the order $o(1)$ and implies that $\text{var}\{(\nabla_{\boldsymbol{\tau}} \boldsymbol{\gamma}_{\boldsymbol{\tau}})'(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_{\boldsymbol{\tau}})\} = (\nabla_{\boldsymbol{\tau}} \boldsymbol{\gamma}_{\boldsymbol{\tau}})' \boldsymbol{\Gamma} (\nabla_{\boldsymbol{\tau}} \boldsymbol{\gamma}_{\boldsymbol{\tau}}) = o(n)$. By the Chebyshev's inequality, we have

$$\sup_{\boldsymbol{\tau}} \left\{ \frac{\|(\nabla_{\boldsymbol{\tau}} \boldsymbol{\gamma}_{\boldsymbol{\tau}})'(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_{\boldsymbol{\tau}})\|}{\|(\nabla_{\boldsymbol{\tau}} \boldsymbol{\gamma}_{\boldsymbol{\tau}})'(\nabla_{\boldsymbol{\tau}} \boldsymbol{\gamma}_{\boldsymbol{\tau}})\|} \right\} \xrightarrow{P} 0,$$

as $n \to \infty$. Thus the convergence of the Newton-Raphson method for $\hat{\boldsymbol{\tau}}$ is ensured. For irregularly grid of data, since $|B(\boldsymbol{s}_i, r) \cap D| = O(r^2)$, it follows that (A1)–(A3) still hold. This completes the proof. □

1.3.3 Proof of Theorem 1

For a given candidate status vector $\boldsymbol{\psi}$ such that $|\boldsymbol{\psi} - \boldsymbol{\delta}| = \boldsymbol{o}^*(n^{1/2})$, we have

$$(\nabla_{\boldsymbol{\lambda}_\psi} \boldsymbol{\theta}_\psi)' \{\boldsymbol{V}(\boldsymbol{\lambda}_\psi, \boldsymbol{\tau}^\dagger)\}^{-1} (\nabla_{\boldsymbol{\lambda}_\psi} \boldsymbol{\theta}_\psi) = (\nabla_{\boldsymbol{\lambda}_\delta} \boldsymbol{\theta}_\delta)' \{\boldsymbol{V}(\boldsymbol{\lambda}_\delta, \boldsymbol{\tau}^\dagger)\}^{-1} (\nabla_{\boldsymbol{\lambda}_\delta} \boldsymbol{\theta}_\delta) + \boldsymbol{o}(n).$$

Thus, $\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{Q}(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau}^\dagger) = \boldsymbol{K}_n\{(\boldsymbol{Y}-\boldsymbol{\theta}_\delta)+(\boldsymbol{\theta}_\delta-\boldsymbol{\theta}_{\boldsymbol{\lambda}_\psi})\} - (\nabla_{\boldsymbol{\lambda}_\delta}\boldsymbol{\theta}_\delta)'\{\boldsymbol{V}(\boldsymbol{\lambda},\boldsymbol{\tau}^\dagger)\}^{-1}(\nabla_{\boldsymbol{\lambda}_\delta}\boldsymbol{\theta}_\delta) + \boldsymbol{o}_p(n)$, where $\boldsymbol{K}_n$ is formed by $\boldsymbol{\lambda}_\psi$ and $\boldsymbol{\tau}^\dagger$. By a central limit theorem (Lin, 2008), we have

$$n^{-1}\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{Q}(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau}^\dagger) \equiv n^{-1}\nabla_{\boldsymbol{\lambda}_\delta}\boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger) \xrightarrow{P} \boldsymbol{I}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger). \tag{A4}$$

Further, by the inverse function theorem, there exists an open ball, $B(\boldsymbol{\lambda}_\delta, r)$ that centers at $\boldsymbol{\lambda}_\delta$ with radius $r > 0$, such that $n^{-1}\boldsymbol{Q}(\boldsymbol{\lambda}_\psi)$ is a one-to-one mapping on the ball with probability tending to one. Also, for this given radius r, there is some $r_0 > 0$ such that

$$B\{n^{-1}\boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger), r_0\} \subseteq n^{-1}\boldsymbol{Q}\left(B(\boldsymbol{\lambda}_\delta, r);\boldsymbol{\tau}^\dagger\right). \tag{A5}$$

By Proposition 1, $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^\dagger + O_p(n^{-1})$ for some $\boldsymbol{\tau}^\dagger$. To establish the consistency of the QL estimate $\hat{\boldsymbol{\lambda}}_\psi^{(m)}$, we observe that

$$n^{-1}\|\boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger) - \boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_\psi^{(m)};\boldsymbol{\tau}^{(m)}\right)\| \le n^{-1}(\|B_1\| + \|B_2\| + \|B_3\|),$$

where $B_1 = \boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger) - \boldsymbol{Q}(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau}^\dagger)$, $B_2 = \boldsymbol{Q}(\boldsymbol{\lambda}_\psi;\boldsymbol{\tau}^\dagger) - \boldsymbol{Q}(\boldsymbol{\lambda}_\psi;\hat{\boldsymbol{\tau}}^{(m)})$, and $B_3 = \boldsymbol{Q}(\boldsymbol{\lambda}_\psi;\hat{\boldsymbol{\tau}}^{(m)}) - \boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_\psi^{(m)};\boldsymbol{\tau}^{(m)}\right)$. Since $B_1 = \left[(\nabla_{\boldsymbol{\lambda}_\delta}\boldsymbol{\theta}_\delta)'\{\boldsymbol{V}(\boldsymbol{\lambda}_\delta,\boldsymbol{\tau}^\dagger)\}^{-1} - (\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{\theta}_\psi)'\{\boldsymbol{V}(\boldsymbol{\lambda}_\psi,\boldsymbol{\tau}^\dagger)\}^{-1}\right](\boldsymbol{Y}-\boldsymbol{\theta}_\delta) + \boldsymbol{o}_p(n^{1/2})$, we have $\sup\{n^{-1}\|B_1\| : \boldsymbol{\lambda} \in \mathbb{R}^{q+2}\} = \boldsymbol{O}_p(n^{-1/2})$ by the central limit theorem. This implies that

$$P(n^{-1}\|B_1\| < r_0/3) \to 0 \ \text{ uniformly as } \ n \to \infty.$$

Next, since $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^\dagger + \boldsymbol{O}_p(n^{-1})$ by Proposition 1, we have $B_2 = (\nabla_{\boldsymbol{\lambda}_\psi}\boldsymbol{\theta}_\psi)'[\{\boldsymbol{V}(\boldsymbol{\lambda}_\psi,\boldsymbol{\tau}^\dagger)\}^{-1} - \{\boldsymbol{V}(\boldsymbol{\lambda}_\psi,\boldsymbol{\tau}^{(m)})\}^{-1}](\boldsymbol{Y}-\boldsymbol{\theta}_\psi) = \boldsymbol{o}_p(1)$ and thus

$$P(\sup_{\boldsymbol{\lambda}} n^{-1}\|B_2\| < r_0/3) \to 0 \ \text{ uniformly as } \ n \to \infty.$$

For B_3, $\boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_\psi^{(m)};\hat{\boldsymbol{\tau}}^{(m)}\right) = \boldsymbol{0}$ almost surely. It can be shown that $n^{-1}\boldsymbol{Q}(\boldsymbol{\lambda}_\psi;\hat{\boldsymbol{\tau}}^{(m)}) = n^{-1}(\nabla_{\boldsymbol{\lambda}_\delta}\boldsymbol{\theta}_\delta)'\boldsymbol{\Lambda}_{\boldsymbol{\lambda}_\delta}(\boldsymbol{Y}-\boldsymbol{\theta}_\delta) + \boldsymbol{o}_p(n^{-1/2})$, which converges to zero uniformly in probability as $n \to \infty$. Thus,

$$P(n^{-1}\|B_3\| < r_0/3) \to 0 \quad \text{uniformly as} \quad n \to \infty.$$

Combining the results above, we have, as $n \to \infty$,

$$P\left(\sup\{n^{-1}\|\boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger) - \boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_\psi^{(m)};\hat{\boldsymbol{\tau}}^{(m)}\right)\| : \boldsymbol{\lambda} \in \mathbb{R}^{q+2}\} < r_0\right) \to 1 \tag{A6}$$

It follows from (A5) and (A6) that,

$$n^{-1}\boldsymbol{Q}\{\hat{\boldsymbol{\lambda}}_\psi^{(m)};\hat{\boldsymbol{\tau}}^{(m)}\} \in B\{n^{-1}\boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger), r_0\} \subseteq n^{-1}\boldsymbol{Q}[B\{(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}^\dagger), r\}], \tag{A7}$$

with probability tending to one as $n \to \infty$. Since r can be arbitrarily small in (A7), the consistency of the QL estimate $\hat{\boldsymbol{\lambda}}_\psi^{(m)}$ follows. $\square$

1.3.4 Proof of Proposition 2

In the three-step computational algorithm, the covariance parameter estimates $\hat{\boldsymbol{\tau}}^{(m)}$ are fixed for all $\boldsymbol{\psi} \in \boldsymbol{\Omega}$. Let $\nabla_{\xi_\psi} \hat{\boldsymbol{\theta}}_\psi^{(m)} = (\nabla_{\xi_\psi} \boldsymbol{\theta}_\psi)\big|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}_\psi^{(m)}}$. It follows from Proposition 1 that

$$\{\nabla_{\xi_\psi} \hat{\boldsymbol{\theta}}_\psi^{(m)}\}'\{\boldsymbol{V}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}, \boldsymbol{\tau}^{(m)})\}^{-1}(\boldsymbol{Y} - \boldsymbol{\theta}_\psi) = \{\nabla_{\xi_\psi} \hat{\boldsymbol{\theta}}_\psi^{(m)}\}'\{\boldsymbol{V}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}, \boldsymbol{\tau}^{\dagger})\}^{-1}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta) + O_p(1). \quad \text{(A8)}$$

Let $\boldsymbol{L}_{\hat{\boldsymbol{\lambda}}_\psi} = \{\nabla_{\xi_\psi} \hat{\boldsymbol{\theta}}_\psi^{(m)}\}'\{\boldsymbol{V}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}, \boldsymbol{\tau}^{\dagger})\}^{-1}$, which is a random function associated with $\hat{\boldsymbol{\lambda}}_\psi$. Note that, given $\hat{\boldsymbol{\lambda}}_\psi$, $\boldsymbol{L}'_{\hat{\boldsymbol{\lambda}}_\psi}$ is an n-dimensional vector with constant elements. So, $\boldsymbol{L}_{\hat{\boldsymbol{\lambda}}_\psi}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta)\Big| \hat{\boldsymbol{\lambda}}_\psi = O_p(n^{1/2})$. Since $\mathrm{var}\{\boldsymbol{L}_{\hat{\boldsymbol{\lambda}}_\psi}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta)\} = E[\mathrm{var}\{\boldsymbol{L}_{\hat{\boldsymbol{\lambda}}_\psi}(\boldsymbol{Y} - \boldsymbol{\theta} - \delta)\}\Big| \hat{\boldsymbol{\lambda}}_\psi] + \mathrm{var}[E\{\boldsymbol{L}_{\hat{\boldsymbol{\lambda}}_\psi}(\boldsymbol{Y} - \boldsymbol{\theta})\}|\hat{\boldsymbol{\lambda}}_\psi]$, we have $\mathrm{var}\{n^{-1/2}\boldsymbol{L}_{\hat{\boldsymbol{\lambda}}_\psi}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta)\} = O(1)$. Thus, (A8) becomes $\{\nabla_{\xi_\psi} \hat{\boldsymbol{\theta}}_\psi^{(m)}\}'\{\boldsymbol{V}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}, \hat{\boldsymbol{\tau}}^{(m)})\}^{-1}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta) = O_p(n^{1/2})$. Also, the QL estimating equation gives $\{\nabla_{\xi_\psi} \hat{\boldsymbol{\theta}}_\psi^{(m)}\}'\{\boldsymbol{V}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}, \hat{\boldsymbol{\tau}}^{(m)})\}^{-1}\{\boldsymbol{Y} - \hat{\boldsymbol{\theta}}_\psi^{(m)}\} = \boldsymbol{0}$ almost surely. Combining these results, we have $\{\nabla_{\xi_\psi} \hat{\boldsymbol{\theta}}_\psi^{(m)}\}'\{\boldsymbol{V}(\hat{\boldsymbol{\lambda}}_\psi^{(m)}, \hat{\boldsymbol{\tau}}^{(m)})\}^{-1}\{\boldsymbol{\theta}_\delta - \hat{\boldsymbol{\theta}}_\psi^{(m)}\} = O_p(n^{1/2})$, which implies that

$$\sum_{j \in \Delta_\psi} \hat{\theta}_{j;\psi}^{(m)} \left[\sum_{i=1}^n \hat{\Lambda}_{j,i;\psi}^{(m)} \{g^{-1}(\beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + \xi_\delta \delta_i) - g^{-1}(\hat{\beta}_0^{(m)} + \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\psi^{(m)} + \hat{\xi}_\psi^{(m)}\psi_i)\} \right] = O_p(n^{1/2}). \quad \text{(A9)}$$

Next, it can be shown that $\hat{\boldsymbol{\beta}}_\psi^{(m)} = \boldsymbol{\beta} + o_p(n^{-1})$. Let

$$t_1 = g^{-1}(\beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + \xi_\delta \delta_i) \quad \text{and} \quad t_2 = g^{-1}\big\{\hat{\beta}_0^{(m)} + \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\psi^{(m)} + \hat{\xi}_\psi^{(m)}\psi_i\big\}.$$

Since $M_0|t_1 - t_2| \geq |g(t_1) - g(t_2)|$ for some $M_0 > 0$ by the mean value theorem, (A9) implies that

$$\sum_{j \in \Delta_\psi} \sum_{i=1}^n \hat{\theta}_{j;\psi}^{(m)} \hat{\Lambda}_{j,i;\psi}^{(m)} \{\beta_0 + \xi_\delta \delta_i - \hat{\beta}_0^{(m)} - \hat{\xi}_\psi^{(m)}\psi_i\} = O_p(n^{1/2}). \quad \text{(A10)}$$

Similarly, for the complement of the jump set, we have

$$\sum_{j \in \Delta_\psi^c} \sum_{i=1}^n \hat{\theta}_{j;\psi^c}^{(m)} \hat{\Lambda}_{j,i;\psi^c}^{(m)} \{(\beta_0 + \xi_\delta) - \xi_\delta \delta_i^c - (\hat{\beta}_0^{(m)} + \hat{\xi}_\psi^{(m)}) + \hat{\xi}_{\psi^c}^{(m)}\psi_i^c\} = O_p(n^{1/2}), \quad \text{(A11)}$$

which completes the proof of Proposition 2. □

1.3.5 Proof of Corollary 2

Since $\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} \xrightarrow{P} \boldsymbol{\lambda}_\delta$ as $n \to \infty$ by Corollary 1, a first-order Taylor expansion implies that

$$\boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}; \boldsymbol{\tau}^{\dagger}\right) = \boldsymbol{Q}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}^{\dagger}) + \left\{ \nabla_{\boldsymbol{\lambda}_\psi} \boldsymbol{Q}(\boldsymbol{\lambda}_\psi; \boldsymbol{\tau}^{\dagger})\Big|_{\boldsymbol{\lambda}_\psi = \hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}} \right\} \{\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} - \boldsymbol{\lambda}_\delta\} + \boldsymbol{o}(\|\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} - \boldsymbol{\lambda}_\delta\|). \quad \text{(A12)}$$

Also, by Proposition 2 and the central limit theorem, we have

$$\boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}; \boldsymbol{\tau}^{\dagger}\right) - \boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}; \boldsymbol{\tau}^{(m)}\right) = O_p(n^{-1/2}). \tag{A13}$$

Combining (A12) and (A13) gives that, with an order of $n^{-1/2}$,

$$\boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}; \hat{\boldsymbol{\tau}}^{(m)}\right) = \boldsymbol{Q}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}^{\dagger}) + \left\{ \nabla_{\boldsymbol{\lambda}_\psi} \boldsymbol{Q}(\boldsymbol{\lambda}_\psi; \boldsymbol{\tau}^{\dagger})\Big|_{\boldsymbol{\lambda}_\psi = \hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}} \right\} \{\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} - \boldsymbol{\lambda}_\delta\} + \boldsymbol{o}(\|\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} - \boldsymbol{\lambda}_\delta\|).$$

Since $\boldsymbol{Q}\left(\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}; \hat{\boldsymbol{\tau}}^{(m)}\right) = \boldsymbol{0}$, the above equation becomes

$$\{\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} - \boldsymbol{\lambda}_\delta\}\{1 + \boldsymbol{o}_p(1)\} = -\left\{ \nabla_{\boldsymbol{\lambda}_\psi} \boldsymbol{Q}(\boldsymbol{\lambda}_\psi; \boldsymbol{\tau}^{\dagger})\Big|_{\boldsymbol{\lambda}_\psi = \hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)}} \right\}^{-1} \boldsymbol{Q}(\boldsymbol{\lambda}_\psi; \boldsymbol{\tau}^{\dagger}), \tag{A14}$$

which implies that

$$n^{1/2}\{\hat{\boldsymbol{\lambda}}_{\psi_0}^{(m)} - \boldsymbol{\lambda}_\delta\} = -n^{1/2}\boldsymbol{Q}(\boldsymbol{\lambda}_\psi; \boldsymbol{\tau}^{\dagger})\boldsymbol{I}^{-1}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}^{\dagger}). \tag{A15}$$

Similarly, we have

$$n^{1/2}\{\hat{\boldsymbol{\lambda}}_{\psi_0^c}^{(m)} - \boldsymbol{\lambda}_{\delta^c}\} = -n^{1/2}\boldsymbol{Q}^*(\boldsymbol{\lambda}_{\delta^c}; \boldsymbol{\tau}^{\dagger})\boldsymbol{I}^{-1}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}^{\dagger}). \tag{A16}$$

Applying the central limit theorem to (A15) and (A16) gives Corollary 2. $\square$

1.3.6 Proof of Theorem 2

If $|\boldsymbol{\psi}-\boldsymbol{\delta}| = \boldsymbol{o}^*(n^{1/2})$, then by Corollary 1, we have $\hat{\xi}_{\psi}^{(m)} = \xi_\delta + O_p(n^{-1/2})$ and $\hat{\xi}_{\psi^c}^{(m)} = -\xi_\delta + O_p(n^{-1/2})$. Recall that $\hat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{\beta} + \boldsymbol{o}_p(n^{-1/2})$. We thus have $\hat{\boldsymbol{\theta}}_{\psi}^{(m)} - \hat{\boldsymbol{\theta}}_{\psi^c}^{(m)} = \boldsymbol{o}_p(n^{-1/2})$, and therefore

$$H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi}^{(m)}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c}^{(m)})\right\} = o_p(n^{-1}).$$

On the other hand, assume that there is one status vector $\boldsymbol{\psi}$ such that $n^{-1/2-c_1}\|\boldsymbol{\delta} - \boldsymbol{\psi}\| \geq M$ for some $M > 0$ and $c_1 > 0$. That is, the number of misclassified sites is $O(n^k)$ for some $k > 1/2$. Then,

$$\sum_{j \in \Delta_\psi} \sum_{i=1}^{n} \hat{\theta}_{j;\psi}^{(m)} \hat{\Lambda}_{j,i;\psi}^{(m)} (\psi_i - \delta_i) = O_p(n^{1/2+c_1}), \tag{A17}$$

and

$$\sum_{j \in \Delta_\psi^c} \sum_{i=1}^{n} \hat{\theta}_{j;\psi^c}^{(m)} \hat{\Lambda}_{j,i;\psi^c}^{(m)} (\psi_i^c - \delta_i^c) = O_p(n^{1/2+c_1}). \tag{A18}$$

Let $c_0 \in (0, c_1)$. By (A10), we have

$$n^{-1/2-c_0} \sup \left\{ \sum_{j\in\Delta_\psi} \sum_{i=1}^{n} \hat{\theta}^{(m)}_{j;\psi} \hat{\Lambda}^{(m)}_{j,i;\psi} \{\beta_0 + \xi_\delta \delta_i - \hat{\beta}^{(m)}_0 - \hat{\xi}^{(m)}_\psi \psi_i\} \right\} \xrightarrow{P} 0, \tag{A19}$$

and by (A11),

$$n^{-1/2-c_0} \sup \left\{ \sum_{j\in\Delta^c_\psi} \sum_{i=1}^{n} \hat{\theta}^{(m)}_{j;\psi^c} \hat{\Lambda}^{(m)}_{j,i;\psi^c} \{(\beta_0 + \xi_\delta) - \xi_\delta \delta^c_i - (\hat{\beta}^{(m)}_0 + \hat{\xi}^{(m)}_\psi) + \hat{\xi}^{(m)}_{\psi^c} \psi^c_i\} \right\} \xrightarrow{P} 0. \tag{A20}$$

One comparison between (A17) and (A19) then gives that $\hat{\xi}^{(m)}_\psi + \hat{\beta}^{(m)}_0 \in (\beta_0, \beta_0 + \xi_\delta) + o_p(n^{-c_1+c_0})$ if $\sum(\psi_i - \delta_i) \geq 0$, and the other comparison between (A18) and (A20) gives that $\hat{\xi}^{(m)}_{\psi^c} + \hat{\beta}^{(m)}_0 \in (\beta_0 - \xi_\delta, \beta_0) + o_p(n^{-c_1+c_0})$ if $\sum(\psi_i - \delta_i) < 0$. This leads to $H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_{\psi^c})\right\} > 0$ almost surely. This proves the "only if" part and thus completes the proof of Theorem 2. □

1.3.7 Proof of Theorem 3

If $\boldsymbol{\delta}^{(m)} = \boldsymbol{\delta} + \boldsymbol{o}^*_p(n^{1/2})$, then by Theorem 2, $H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_{\delta^{(m)}}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_{\delta^{(m)c}})\right\} \xrightarrow{P} 0 = \min_{\boldsymbol{\psi}\in\boldsymbol{\Omega}} H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\right\}$. So, $\boldsymbol{\delta}^{(m)} = \arg\min_{\boldsymbol{\psi}\in\boldsymbol{\Omega}} H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\right\}$. To show the "only if" part, let $\boldsymbol{\delta}^* = \boldsymbol{\delta} + \boldsymbol{o}^*(n^{1/2}) \in \boldsymbol{\Omega}$ and let $\boldsymbol{\delta}^*_c = 1 - \boldsymbol{\delta}^*$. Then, we have

$$H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_{\delta^{(m)}}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_{\delta^{(m)c}})\right\} = \min_{\boldsymbol{\psi}\in\boldsymbol{\Omega}} H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_\psi), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\psi^c})\right\} \leq H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\delta^*}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\delta^{*c}})\right\}.$$

Since by Theorem 2, $H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\delta^*}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\delta^{*c}})\right\} + o_p(n^{-1})$, we have

$$H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_{\delta^{(m)}}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}^{(m)}_{\delta^{(m)c}})\right\} = H\left\{\boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\delta^*}), \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}}_{\delta^{*c}})\right\} + o_p(n^{-1}).$$

Note that the homogeneity measure is a quadratic function with a unique minimum value at $\boldsymbol{\delta}$, and we thus have $\boldsymbol{\delta}^{(m)} \equiv \boldsymbol{\delta}^* = \boldsymbol{\delta} + \boldsymbol{o}^*_p(n^{1/2})$. This completes the proof. □

1.3.8 Proof of Theorem 4

Note that $\sum_{j\in\Delta_\psi} \sum_{i=1}^n \hat{\theta}^{(m)}_{j;\psi} \hat{\Lambda}^{(m)}_{j,i;\psi} (\xi_\delta \delta_i + \psi_i) = O_p(n)$, and therefore when $\delta_i = 0$ for all i, we have $\hat{\xi}^{(m)}_\psi = O_p(n^{-1/2})$. Also, by (A11), $\hat{\xi}^{(m)}_{\psi^c} = O_p(n^{-1/2})$, and thus $\hat{\xi}^{(m)}_\psi + \hat{\xi}^{(m)}_{\psi^c} = O_p(n^{-1/2})$ when $\boldsymbol{\delta} \equiv \boldsymbol{0}$. This implies part (a) of Theorem 4. Furthermore, since $(\hat{\xi}^{(m)}_\psi + \hat{\xi}^{(m)}_{\psi^c}) \xrightarrow{P} 0$ as $n \to \infty$, a Taylor expansion for $\boldsymbol{Q}(\hat{\boldsymbol{\lambda}}^{(m)}_\psi; \hat{\boldsymbol{\tau}}^{(m)}) + \boldsymbol{Q}^*(\hat{\boldsymbol{\lambda}}^{(m)}_{\bar{\delta}^c}; \boldsymbol{\tau}^{(m)})$ with respect to $\xi_\psi = 0$ then gives $n^{1/2}\{\hat{\xi}^{(m)}_\psi + \hat{\xi}^{(m)}_{\psi^c}\} = n^{-1}\sigma^{-1}_\psi \left[\boldsymbol{U}'_\psi(\boldsymbol{Y} - \boldsymbol{\theta}_\delta)\right]$. Applying the central limit theorem leads to the asymptotic normal distribution of Z_ψ. □

1.3.9 Proof of Theorem 5

For convenience, let $\boldsymbol{D} = \nabla_{\lambda_\delta}\boldsymbol{\theta}_\delta$ denote an $n \times (q+2)$ matrix. Also, we use $\boldsymbol{V}^{-1}$ to denote an inverse matrix of the covariance matrix $\boldsymbol{V}$. We define a matrix operator by $(A_1, \ldots, A_k) \odot B = (A_1 B, \ldots, A_k B)$, where $A_1, \ldots, A_k$ and B are matrices. Then

$$\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}) = (\partial \boldsymbol{D}'/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{D}'/\partial \boldsymbol{\beta}_q, \partial \boldsymbol{D}'/\partial \xi) \odot \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta) + \boldsymbol{D}' \odot (\partial \boldsymbol{V}^{-1}/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{V}^{-1}/\partial \boldsymbol{\beta}_q, \partial \boldsymbol{V}^{-1}/\partial \xi)(\boldsymbol{Y} - \boldsymbol{\theta}_\delta) - \boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D}. \tag{A21}$$

By Assumptions 1-2, $(\partial \boldsymbol{D}'/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{D}'/\partial \xi) \odot \boldsymbol{V}^{-1}$ and $\boldsymbol{D}' \odot (\partial \boldsymbol{V}^{-1}/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{V}^{-1}/\partial \xi)$ are bounded vectors. A central limit theorem for random fields (Guyon, 1995) implies that a multivariate normal random variable exists $\boldsymbol{W}$ such that

$$n^{-1/2}\{(\partial \boldsymbol{D}'/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{D}'/\partial \xi) \odot \boldsymbol{V}^{-1} + \boldsymbol{D}' \odot (\partial \boldsymbol{V}^{-1}/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{V}^{-1}/\partial \xi)\}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta) = \boldsymbol{W} + \boldsymbol{o}_p(1),$$

which implies that

$$\{(\partial \boldsymbol{D}'/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{D}'/\partial \xi) \odot \boldsymbol{V}^{-1} + \boldsymbol{D}' \odot (\partial \boldsymbol{V}^{-1}/\partial \boldsymbol{\beta}_0, \ldots, \partial \boldsymbol{V}^{-1}/\partial \xi)\}(\boldsymbol{Y} - \boldsymbol{\theta}_\delta) = \boldsymbol{O}_p(n^{1/2}).$$

Plugging the above result into (A21) thus gives

$$n^{-1}\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}) = \boldsymbol{O}_p(n^{-1/2}) - n^{-1}\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D}, \tag{A22}$$

which converges in probability to a positive-definite matrix $I_0(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$ as $n \to \infty$ by Assumption 3. Note that this result also implies that $n^{-1}\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$ is uniformly bounded over n in probability. □

1.3.10 Proof of Corollary 3

For two parameter vectors $\boldsymbol{\lambda}_\delta$ and $\boldsymbol{\lambda}_\delta^*$, it follows from (A22) that $n^{-1}\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}) = \boldsymbol{O}_p(n^{-1/2}) - n^{-1}\boldsymbol{D}_1'\boldsymbol{V}_1^{-1}\boldsymbol{D}_1$ and $n^{-1}\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta^*; \boldsymbol{\tau}) = \boldsymbol{O}_p(n^{-1/2}) - n^{-1}\boldsymbol{D}_2'\boldsymbol{V}_2^{-1}\boldsymbol{D}_2$, where $\boldsymbol{D}_1 \equiv \nabla_{\lambda_\delta}\boldsymbol{\theta}_\delta$, $\boldsymbol{D}_2 \equiv \nabla_{\lambda_\delta^*}\boldsymbol{\theta}_\delta$, $\boldsymbol{V}_1^{-1} \equiv \boldsymbol{V}_{\lambda_\delta}^{-1}$, and $\boldsymbol{V}_2^{-1} \equiv \boldsymbol{V}_{\lambda_\delta^*}^{-1}$. Then, $n^{-1}\|\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau}) - \dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta^*; \boldsymbol{\tau})\| \leq n^{-1}\{\|(\boldsymbol{D}_1 - \boldsymbol{D}_2)'\boldsymbol{V}_1^{-1}\boldsymbol{D}_1\| + \boldsymbol{D}_2'(\boldsymbol{V}_1 - \boldsymbol{V}_2)^{-1}\boldsymbol{D}_1\| + \|\boldsymbol{D}_2'\boldsymbol{V}_2^{-1}(\boldsymbol{D}_1 - \boldsymbol{D}_2)'\|\} + \boldsymbol{O}_p(n^{-1/2})$. Recall that $\boldsymbol{D} = \partial g^{-1}(\beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \xi_\delta\boldsymbol{\delta})/\partial(\beta_0, \boldsymbol{\beta}, \xi)$. It thus follows from Assumptions 1 and 2 that $\|\boldsymbol{D}\|$ and $\boldsymbol{V}\|$ are bounded for any given parameters $\boldsymbol{\lambda}_\delta$. Since $\boldsymbol{D}'\boldsymbol{V}\boldsymbol{D}$ is uniformly bounded by Proposition 3, we can find a constant $l^* > 0$ such that $n^{-1}\|(\boldsymbol{D}_1 - \boldsymbol{D}_2)'\boldsymbol{V}_1^{-1}\boldsymbol{D}_1\|$, $n^{-1}\|\boldsymbol{D}_2'(\boldsymbol{V}_1 - \boldsymbol{V}_2)^{-1}\boldsymbol{D}_1\|$, and $n^{-1}\|\boldsymbol{D}_2'\boldsymbol{V}_2^{-1}(\boldsymbol{D}_1 - \boldsymbol{D}_2)'\|$ are all uniformly bounded by $l^*\|\boldsymbol{\lambda}_\delta - \boldsymbol{\lambda}_\delta^*\|$ over n. This completes the proof for the Lipschitz condition of $\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta; \boldsymbol{\tau})$. □

1.3.11 Proof of Corollary 4

By proof of Proposition 3, we have, for any $\boldsymbol{\lambda}_\delta$, $\boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau}) = \boldsymbol{O}_p(n^{1/2})$ and $\{\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta);\boldsymbol{\tau})\}^{-1} = \boldsymbol{O}_p(n^{-1})$, which implies that $\|\{\dot{\boldsymbol{Q}}(\boldsymbol{\lambda}_\delta);\boldsymbol{\tau})\}^{-1}\boldsymbol{Q}(\boldsymbol{\lambda}_\delta;\boldsymbol{\tau})\| = O(n^{-1/2}) \to 0$ as $n \to \infty$. This implies that, as $n \to \infty$, the Newton-Kantorovich method (Argyros, 2008) for the QL function converges under any initial values. □

1.3.12 Proof of Corollary 5

First, recall that the initial estimate $\hat{\lambda}_\delta$ is computed under an independence assumption. By McCullagh (1983), the QL estimate $\hat{\lambda}_\delta$ is consistent, and a Newton iteration for the QL function converges in the initial step. Also, note that Assumptions 1-3 satisfy required conditions for linear convergence shown in Theorem 1 of Jiang et al. (2007). The iterative estimation procedure of Algorithm 1 thus converges in probability as $n \to \infty$. □

2 Extra Simulation Studies

2.1 Studies for Convergency and Consistency

We conducted a simulation study to evaluate the finite-sample performance of the homogeneity measure. An $l \times l$ grid was considered for $l = 12$, 18, or 24 and thus the sample sizes are $n = 12^2, 18^2$, or 24^2. For the ith cell on the grid, we let the upper-left corner point $\boldsymbol{s}_i = (r-1, c-1)'$ be the representative site for the cell on the rth row (from top to bottom) and the cth column (from left to right).

We considered two cases for the jump sets, as shown in Figure A1 for 24×24 grids. The first case (Figure A1(a)) has a single jump set $\mathcal{C}_1$ in the top one-third of the spatial domain with cardinality $|\mathcal{C}_1| = n/3$. The second case (Figure A1(b)) also has a single jump set $\mathcal{C}_2$ in the top $72/l$ rows; that is the jump set size is fixed at $|\mathcal{C}_2| \equiv 72$ for different l (or n). Let $\boldsymbol{\delta}^{\mathcal{C}_t} = \left(\delta_1^{\mathcal{C}_t}, \ldots, \delta_n^{\mathcal{C}_t}\right)'$ denote the status vector such that $\delta_i^{\mathcal{C}_t} = I[\boldsymbol{s}_i \in \mathcal{C}_t]$ for $t = 1, 2$, and let ξ_t denote the jump coefficient associated with $\boldsymbol{\delta}^{\mathcal{C}_t}$. We used horizontal lines to define candidate jump sets. Let $\mathcal{L}_k$ denote the horizontal line $r = k - 1$, for $k = 1, \ldots, l$ and the rth row. Define the kth candidate status vector $\boldsymbol{\psi}_k = (\psi_k(\boldsymbol{s}_1), \ldots, \psi_k(\boldsymbol{s}_n))'$ where $\psi_k(\boldsymbol{s}_i) = I[\boldsymbol{s}_i$ is above or on the line $\mathcal{L}_k]$. Thus, the collection of all candidate status vectors is $\boldsymbol{\Omega}^l = \{\boldsymbol{\psi}_k : k = 1, \ldots, l\}$ for the $l \times l$ grid. Note that the number of candidate status vectors $|\boldsymbol{\Omega}^l|$ increases with the sample size.

We generated random effects ϵ_i from a Gaussian distribution with mean zero, variance σ^2, and correlation $\rho_{i,j} = \rho^{\|\boldsymbol{s}_i - \boldsymbol{s}_j\|}$ for $i, j = 1, \ldots, n$. The covariate x_i was generated from a standard normal distribution. Given ϵ_i and x_i, we considered two types of response variables. First, we generated continuous responses by a Gaussian model $Y_i^{\mathrm{G}} = \beta_0 + \beta_1 x_i + \xi_t \delta_i^{\mathcal{C}_t} + \epsilon_i$, $t = 1, 2$. Next, count responses were generated by a Poisson model with the conditional mean $\theta_{i;\epsilon_i}^{\mathrm{P}} = \exp\left(\beta_0 + \beta_1 x_i + \xi_t \delta_i^{\mathcal{C}_t} + \epsilon_i\right)$, for $t = 1, 2$. By the moment generating functions, the unconditional marginal means are $\theta_i^{\mathrm{P}} = \exp\left(0.5\sigma^2 + \beta_0 + \beta_1 x_i + \xi_t \delta_i^{\mathcal{C}_t}\right)$, for $t = 1, 2$. The jump coefficients were $\xi_1 = \xi_2 = 1$ for the single jump set. We also let $\beta_0 = 0$, $\beta_1 = 1$, $\sigma^2 = 0.5$, and $\rho = 0.5$. For each setting, we simulated 500 data sets. In the simulation, the covariance function for the continuous data is $\mathrm{Cov}(Y_i^{\mathrm{G}}, Y_j^{\mathrm{G}}) = \sigma^2 \rho_{i,j}$, and for the count responses, the covariance structure can be computed as $\mathrm{var}(Y_i^{\mathrm{P}}) = \theta_i^{\mathrm{P}} + \left(\theta_i^{\mathrm{P}}\right)^2 \{\exp(\sigma^2) - 1\}$ and $\mathrm{cov}(Y_i^{\mathrm{P}}, Y_j^{\mathrm{P}}) = \theta_i^{\mathrm{P}} \theta_j^{\mathrm{P}} \{\exp(\sigma^2 \rho_{i,j}) - 1\}$ for $i \neq j$. We then use the homogeneity measure to identify the jump sets in various simulation settings.

Table A1 shows identification rates for the single jump set cases $\mathcal{C}_1$ and $\mathcal{C}_2$ at various sample sizes and iteration numbers $m = 1$ and M, where M denotes a total number of iterations for convergency. In Table A1, we find that for all the iteration numbers, the identification rate of the true jump set increases as the sample size increases. These simulation results support the theory that the homogeneity measure is consistent. Additionally, Table A1 indicates that the homogeneity measure can achieve convergency quite quickly for the Gaussian and Poisson responses, as can be seen that the identification results for iteration numbers 1 and M are pretty close. Specifically, for the Gaussian responses, $M = 1$ in about 95% of simulation runs, while for the Poisson responses, $M = 4$ in about 90% of simulation runs. Nevertheless, we also find that the identification performance would be affected by linearity of the link function. For example, for the 12×12, 18×18, and 24×24 grids, the identification rates for the Gaussian responses with an identity link are around 90%, 95%, and 99%, respectively, while those decrease to 75%, 85%, and 90%, respectively, for the Poisson responses with a log link. The result that the homogeneity measure has better performance in the Gaussian distribution than that in the Poisson distribution is not unexpected. As shown in (2.5), the homogeneity measure is derived from a linear approximation for the log-QL ratio, and therefore performance of the proposed method would lean on linearity of the link function.

For estimation results of the model parameters, Table A2 shows sample means and sample variances for HLR estimates of β_1 and ξ_1 in the Gaussian and Poisson models. (The estimation result for ξ_2 is similar to that of ξ_1, so we present only the estimation result for ξ_1.) The estimates of β_1 and ξ_1 are quite unbiased on the 18×18 and 24×24 grids for both the Gaussian and Poisson models. Table A2 also shows consistency for the HLR estimates of β_1 and ξ_1, as the sample variance decreases when the sample size increases. Nevertheless, when comparing sample variances between different models, we find that the homogeneity measure also has better performance in the linear model than that in the log-linear model. This pattern is similar to the identification result in Table A1.

2.2 Validation Studies for the Weighted Least Squares Errors

To evaluate whether the weighted least squares errors used in Section 5 is suitable for model selection in the data analysis, we conduct a simulation study based on a geographic structure similar to the analysis result for the Midwest data. In this simulation study, four jump sets, $\Delta_1, \ldots, \Delta_4$, are chosen with $|\Delta_1| = 5$, $|\Delta_2| = 20$, $|\Delta_3| = 150$, and $|\Delta_4| = 160$ (Figure A3). The jump sets include about 64% of the total counties. We then generate data by a procedure similar to that in Section 4. Specifically, for each simulation run, we generate $\theta_i^* = \beta_0 + \sum_{q=1}^{5} \beta_q x_{q,i} + \sum_{k=1}^{4} \xi_k \delta_{k,i} + \epsilon_i$, where $\delta_{k,i} = I[\boldsymbol{s}_i \in \Delta_k]$ denotes the status variable associated with Δ_k, $k = 1, \ldots, 4$, $i = 1, \ldots, 535$. The spatial noise ϵ_i is generated from a multivariate Gaussian distribution with mean zero, $\sigma = 0.01$, and correlation $\mathrm{corr}(\epsilon_i, \epsilon_j) = 0.6 \exp(-0.5 \|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$. Let $Y_i^* = \exp(\theta_i^*)/\{1 + \exp(\theta_i^*)\}$, $i = 1, \ldots, 535$. We follow the transformation process in the data analysis to generate responses by $Y_i = \log\{Y_i^*/(1 - Y_i^*)\}$. The regression coefficients are set to be $\beta_0 = -1.5$, $\beta_1 = 0.8$, $\beta_2 = -0.2$, $\beta_3 = 0.7$, $\beta_4 = -4.5$, and $\beta_5 = 0.9$, and the jump coefficients to be $\xi_1 = 1.6$, $\xi_2 = 1.2$, $\xi_3 = 0.8$, and $\xi = 0.5$. We conduct a total of 200 simulation runs to evaluate the performance.

Nevertheless, in this simulation setting, the HM-QL method tends to present only two identified jump sets in most cases, in which Δ_2, Δ_3, and Δ_4 are combined as a big jump set by the HM-QL method. Therefore, we show the simulation result only for the HLR method. Tables A3 and A4 show the identification rate and estimation result, respectively, for the homogeneity measure. In Table A4, we also compare average values of weighted least squares errors over the 200 replicates for the regression model without jump sets and the jump set model with covariates and four identified jump sets. As can be seen from Table A3, when numbers of counties in the jump set increase, the HLR method can still identify the jump sets $\Delta_1, \ldots, \Delta_4$ very accurately. Also, for the estimation result by the HLR method shown in Table A4, we find that most estimated coefficients are unbiased and significant. On the other hand, most estimated coefficients in the regression model without the jump sets are seriously biased, indicating that when jump sets exist, ignoring the jump set effects would lead to incorrect statistical inference. Table A4 also shows that an average value of WLS errors in the jump set model is about 0.5, while that in the regression model is about 50. With a comparison to an F-distribution with degrees of freedom at 529 and 525, we find that the difference in means of WLS errors between the two models is extremely significant. This may indicate that using the WLS errors for model selection in the data analysis is suitable.

3 More Discussions

In the case study, we have found that the state-level jump sets are not significant except for Michigan, suggesting a smaller scale for the spatial patterns of jump sets. Therefore, we have analyzed a county-level socio-economic data set where jump sets are present to show interesting patterns additional to the role of race and industrial structure in poverty prior to the war on poverty in the 1960's. The analysis resulted in 392 out of 535 counties (about 73%) that belonged to a jump set; that is, their poverty rates are not adequately explained by the known race and industrial structures. In contrast, the remaining 143 counties whose poverty rates can be fully explained by the known race and industrial structures were among the wealthiest counties with the lower poverty rates. In addition, the jump-set with the highest poverty rates ($\bar{\Delta}_1$) and the jump-set with the lower poverty rates ($\bar{\Delta}_4$) took place both in Illinois. The state of Minnesota had more moderate poverty rates with more middle-class counties whose poverty rate can't be adequately explained by the racial and industrial covariates.

We have also conduced simulation studies in Section 4 to compare the proposed approach with an existing HM-QL method. The numerical examples have indicated that the HLR method is more powerful than the HM-QL method in identifying jump sets for irregularly shaped geographical cells when many covariates are involved. This is plausible because the HLR method compares the difference between the marginal mean functions, while the HM-QL method relies on variable selection in nested models. In search of varying shapes of jump sets mixed with different types of explanatory variables, our proposed method tends to be more robust in dealing with misspecified models, while the HM-QL method is faster in computation and easier to maintain probability of Type I errors within a desired level. Nevertheless, although the proposed HLR method can apply to generalized linear models, its identification ability for jump sets has the best shot in linear models, as can be seen from the derivation process for the homogeneity measure.

References

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

Fingleton, B. (1983). Independence, Stationarity, Categorical Spatial Data and the Chi-Squared Test. *Environment and Planning A* **15**, 483-499.

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

Jiang, J., Luan, Y., Wang, Y.-G. (2007). Iterivative estimating equation: linear convergence and asymptotic properties. *Annals of Statistics* **35**, 2233-2260.

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

McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statististics* **11**, 59-67.

Figure A1. Plots for two types of jump sets in a 24×24 grid of cells. (a) $\mathcal{C}_1$: the proportion of the area of one jump set is fixed at 1/3 (grey). (b) $\mathcal{C}_2$: the area of one jump set is fixed at 4×12 (grey).

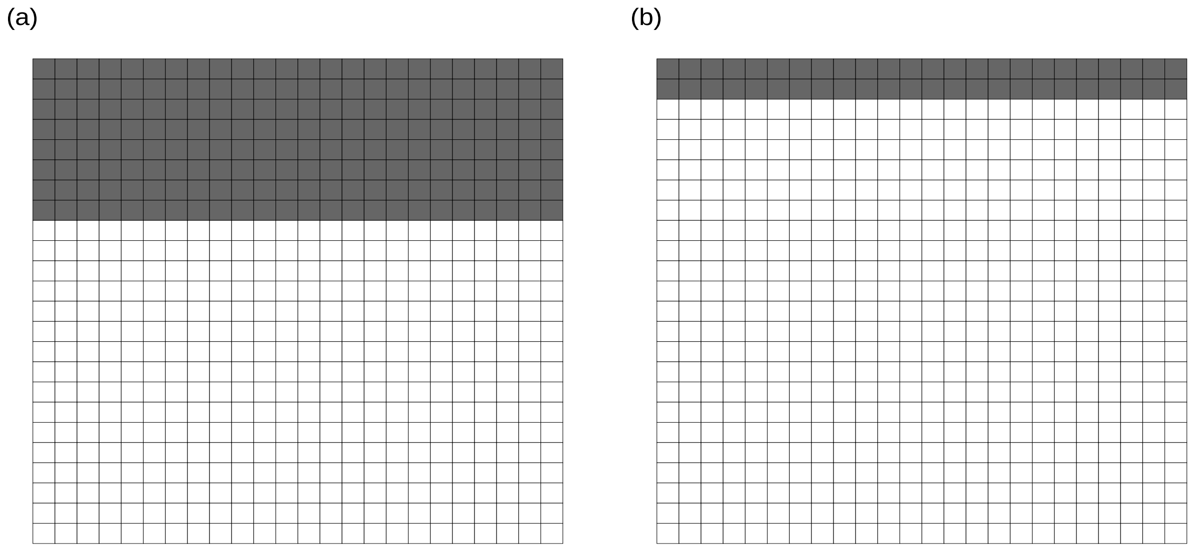

Figure A2. A map for counties in the true jump sets, $\Delta_1, \ldots, \Delta_3$, in the simulation.

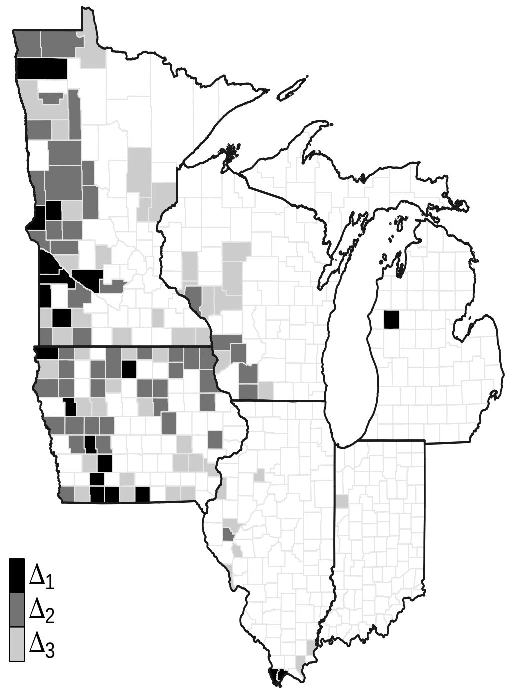

Figure A3. A map for counties in the true jump sets, $\Delta_1, \ldots, \Delta_4$, in the simulation of the Supplementary Material.

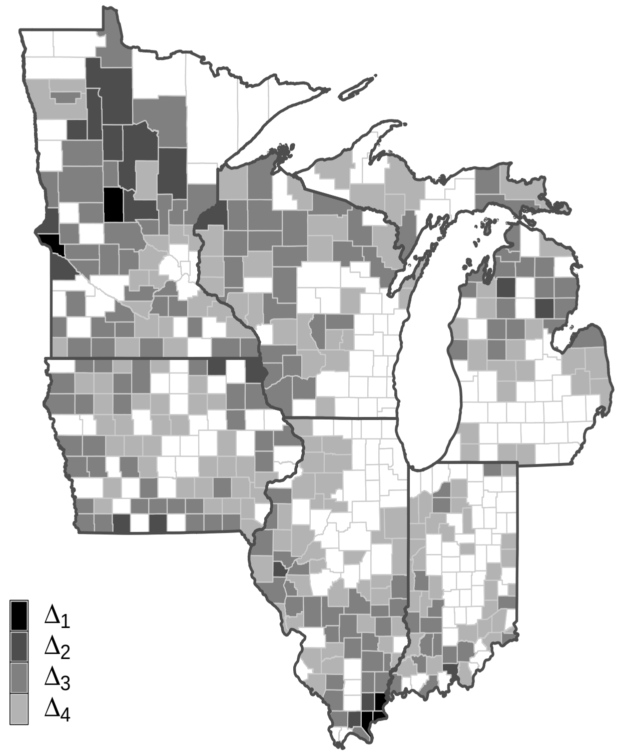

Table A1. For a single jump-set ($\mathcal{C}_1$ and $\mathcal{C}_2$), numbers of simulation times that the true sets are identified by the homogeneity measure at iteration $m = 1$ or M (final iteration), based on 500 replicates for Gaussian and Poisson models on the 12×12, 18×18, 24×24 grids.

	Gaussian		Poisson	
Iteration	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$
		12×12		
1	459	453	381	378
M	460	454	385	381
		18×18		
1	487	481	437	414
M	488	482	440	418
		24×24		
1	496	488	468	439
M	496	488	471	441

Table A2. Sample means and sample variances (in parenthesis) of the regression coefficient β_1 and the jump coefficients ξ_1, based on 500 replicates in the Gaussian and Poisson models on the 12×12, 18×18, 24×24 grids for one jump set.

	Gaussian		Poisson	
	β_1	ξ_1	β_1	ξ_1
	1	1	1	1
12×12	1.00	0.94	1.04	1.05
	(0.12)	(0.52)	(0.50)	(0.74)
18×18	1.00	0.99	1.00	1.00
	(0.03)	(0.19)	(0.07)	(0.36)
24×24	1.00	1.00	1.00	0.99
	(0.02)	(0.15)	(0.06)	(0.27)

Table A3. For the Midwest data structure, the average numbers of counties in the true jump set Δ_i, $i = 1, \ldots, 4$, that were classified into $\bar{\Delta}_j$ by HLR method. The numbers of counties in each jump set are $|\Delta_1| = 5$, $|\Delta_2| = 20$, $|\Delta_3| = 150$, and $|\Delta_4| = 160$. The simulation result is based on 200 replicates.

True	$\bar{\Delta}_1$	$\bar{\Delta}_2$	$\bar{\Delta}_3$	$\bar{\Delta}_4$
Δ_1	5.0	0.0	0.0	0.0
Δ_2	0.0	20.0	0.0	0.0
Δ_3	0.0	0.0	150.0	0.0
Δ_4	0.0	0.0	0.0	158.8

Table A4. Estimation results for the simulation based on the poverty data. Sample means (Mean) and sample variances× 100 (Var) are listed for estimates in the regression model (without jumpsets) and the final jumpset model with covariates $x_1, \ldots, x_5$, and four jumpsets $\bar{\Delta}_1, \ldots, \bar{\Delta}_4$. Average numbers of weighted least squares (WLS) errors are also reported. Simulation results are based on 200 replicates.

		No Jumpsets		With Jumpsets	
	True	Mean	Var	Mean	Var
β_0	-1.5	-0.8	0.01	-1.5	0.009
β_1	0.8	1.0	0.02	0.8	0.02
β_2	-0.2	-1.3	0.02	-0.2	0.02
β_3	0.7	1.8	0.05	0.73	0.2
β_4	-4.5	-6.7	1.0	-4.6	4.0
β_5	0.9	-0.3	0.03	0.85	1.0
ξ_1	1.6			1.6	0.009
ξ_2	1.2			1.2	0.007
ξ_3	0.8			0.8	0.004
ξ_4	0.5			0.5	0.003
WLS error		50.5		0.45	

Table A5. Parameter estimates with standard errors (in parentheses) and p-values for the poverty case study based on the regression model (without jumpsets) and final jumpset model.

	No Jumpsets		With Jumpsets	
explanatory variable	Estimate	p-value	Estimate	p-value
Intercept	-1.6(0.12)	<0.01	-1.8(0.08)	<0.01
x_1	2.1(0.14)	<0.01	1.1(0.09)	<0.01
x_2	-0.7(0.14)	<0.01	-0.3(0.09)	<0.01
x_3	1.3(0.34)	<0.01	1.0(0.22)	<0.01
x_4	-6.9(0.10)	<0.01	-4.7(0.66)	<0.01
x_5	2.3(0.27)	<0.01	1.3(0.18)	<0.01
$\bar{\Delta}_1$			1.2(0.06)	<0.01
$\bar{\Delta}_2$			1.0(0.03)	<0.01
$\bar{\Delta}_3$			0.7(0.03)	<0.01
$\bar{\Delta}_4$			0.6(0.02)	<0.01
$\bar{\Delta}_5$			0.3(0.02)	<0.01
WLS error	494		125	

Computer Code for Data Analysis

```
library("gstat")
# Input Midwest data
dataset <- read.csv("upMidWestpov_Iowa_cluster_names_race_regime.csv")
dataset.names <- names(dataset)
dataset60 <-
  dataset[,c("pindpov60","logitindpov60","pag60","pman60","pserve60","pfire60","pblk60","x","y
  ")]
y_s <- dataset60$pindpov60; y <- log(y_s/(1-y_s))
covariate <- dataset60[,3:7]
data.A <- dataset60
n <- nrow(data.A)
X <- cbind(1,covariate); R <- y
xy <- data.frame(cbind(x=data.A$x,y=data.A$y))

# Euclidean distance matrix
disL2 <- dist(xy, method="euclidean")
dist2full <- function(dis){
  n <- attr(dis, "Size") ; full <- matrix(0, n, n)
  full[lower.tri(full)] <- dis
  return(full + t(full)) }
L2 <- dist2full(disL2)

# Creation of candidate jump sets
range.seq <- seq(from=-2.8,to=0,by=0.01)
m <- length(range.seq); W <- matrix(NA, n, m)
for (i in 1:(m)){
  W[,i] <- ifelse(y> range.seq[i], 1, 0) }
W_c <- 1-W # complementary of candidate jump sets

# Estimation of Correlation
vgm.est <- function(resi.in, loc.in, dist.in, yvar,cut.off) {
  r <- resi.in # residuals
  resid.loc <- as.data.frame(cbind(loc.in, r=r))
  gamma.vgm=variogram(r ~ x + y, locations = ~x + y,
                      cutoff=sqrt(length(r))*sqrt(2)*2/3, data = resid.loc)
  if(cut.off==T){
```

```
      gamma.vgm=variogram(r ~ x + y, locations = ~x + y, data = resid.loc) }
    fit.vgm <- tryCatch(
    {fit.variogram( gamma.vgm,model=vgm(psill=mean(gamma.vgm[,"gamma"]),
    model="Exp",range=which.max(abs(diff(gamma.vgm[1:5,"gamma"])))+1,
    Nugget=mean(gamma.vgm[, "gamma"])/10),fit.method=2)},warning = function(w)
    {rho.key=1;print(rho.key)} )
  return(corr) }

# Quasi-likelihood Estimation
QLE=function(beta.in,X.in,R.in,n_i,Corr,sigma2,fix_par) {
    ly <- length(R.in)        # sample size
    lb <- length(beta.in) # number of parameters
    beta.in0 <- beta.in
    X <- cbind(X.in)
    fix_par <- fix_par
    V1      <- Corr              # first component of the covariance matrix of Y
    invI <- diag(0,lb) # cov(betahat): inverse of information matrix
    iter <- 0
    repeat{
       res=tryCatch({
          iter <- iter+1
          mu <- as.vector(X%*%beta.in +sum(fix_par)) # E(Y)
          D    <- X    # derivative matrix, depending on the link function
          V    <- V1*sigma2
          detV <- det(V)
          invV <- solve(V,LINPACK=T,tol=1e-400)
          I <- t(D)%*%invV%*%D    #information matrix for the QL estimates
          detI <- det(I)
          invI <- solve(I,LINPACK=T,tol=1e-400) # Cov(betahat)
          beta1 <- beta.in
          beta.in <- beta.in+(invI%*%t(D)%*%invV%*%(R.in-mu)) # Newton-Raphson
    list(iteration=iter,betahat=beta.in,covbetahat=invI ) }

# Merge jump sets
Merg <- function(Max.T.all,W) {
    N <- length(Max.T.all)
    U <- array(0,c(nrow(W),n))
    u[,1] <- W[,sort(Max.T.all,decreasing = T)[1]]
```

```
  for(i in 2:n){
    v <- rep(0,nrow(W))
    v[setdiff(which(W[,sort(Max.T.all,decreasing = T)[i]]==1),
    which(W[,sort(Max.T.all,decreasing = T)[i-1]]==1))] <- 1
    u[,i] <- v }
  u <- as.matrix(u)   }

# Identification procedure for single jump set
X1 <- as.matrix(covariate)
X1_c <-as.matrix(covariate)
QLR.S=sigma2=sigma2_xi =Z_w=p= array(NA,c(1,m))
par.out <- array(NA,c(7,m))
par.out_c <- array(NA,c(7,m))
for (k in 1:m){
# Estimation for main models
  R.in <- R
  X.in <- cbind(1,X1,W[,k])
  model1=lm(R.in~0+X.in[,-1],offset= rep((mean(R)),n)) # initial values
  par.in = c((mean(R)),model1$coefficients)
  yvar.in = as.vector(var(R.in-(X.in%*%par.in)))
  qle.model1 <-
    QLE(beta.in= par.in[-1],X.in=X.in[,-1],R.in=R.in,n_i=1,
      Corr=diag(1,n),sigma2=yvar.in,fix_par=par.in[1])
 # Estimation for complementary models
  R.in_c <- R
  X.in_c <- cbind(1,X1_c,W_c[,k])
  model2 <- lm(R.in_c~0+X.in_c[,-1],offset= rep(sum(par.in[-2:-6]),n))
  par.in_c    <-    c((mean(R)),model2$coefficients)
  yvar.in_c <- as.vector(var(R.in_c-(X.in_c%*%par.in_c)))
  qle.model2 <-
    QLE(beta.in= par.in_c[-1],X.in=X.in_c[,-1],R.in=R.in_c,n_i=1,
      Corr=diag(1,n),sigma2=yvar.in_c,fix_par=par.in[-2:-6])

  tryCatch(
    {sigma2[k]=qle.model1$Sigma_u^2+qle.model2$Sigma_u^2},error=function(e){})
  par.out[,k]      <- c((mean(R)),qle.model1$betahat)
  par.out_c[,k] <-    c((mean(R)),qle.model2$betahat)
  theta.in      <- X.in%*%par.in
```

```
    theta.in_c <- X.in_c%*%par.in_c
    Vinv      <- diag(1,n)*yvar.in
    Vinv_c <- diag(1,n)*yvar.in_c
    QLR.S[k] <-
    ((2*n)^(-1))*t(theta.in-theta.in_c)%*%(Vinv+Vinv_c)%*% (theta.in-theta.in_c)

# Z-test for significance
    U <- solve(t(W[,k])%*%Vinv%*%W[,k],tol=1e-300)%*%W[,k] +
            solve(t(W_c[,k])%*%Vinv_c%*%W_c[,k],tol=1e-300)%*%W_c[,k]
    sigma2_xi[k] <- U %*%Vinv%*%t(U)
    Z_w[k] <- (par.out[7,k]-par.out_c[7,k])/sqrt(sigma2_xi[k])
    p[k]      <- (1-pnorm(Z_w[k]))*2    }

Max.T= which.min(QLR.S) # estimated jump set

# Search for multiple jump sets
Count <- 0; Max.T2 <- -2
Max.T.all <- NULL; Max.T2.all <- NULL
QLR.S=sigma2=sigma2_xi =Z_w=p= array(NA,c(1,m))
while(Max.T2!=Max.T){
    count <- count+1
    if(count!=1){Max.T=Max.T2}
    R.in <- R
    X.in <- cbind(1,X1,W[,Max.T])
    model1 <- lm(R.in~X.in[,-1])
    par.in <-    model1$coefficients
    yvar.in <-    as.vector(var(R.in-(X.in%*%par.in)))

    cov.vgm <- vgm.est(resi.in=(R.in-(X.in%*%par.in)),loc.in=xy,dist.in=L2,
     yvar=yvar.in,cut.off=F)
    Corr.out<- cov.vgm
    qle.model1 <- QLE(beta.in=par.in[-1],X.in=X.in,R.in=R.in,n_i=1,
       Corr=Corr.out,sigma2=yvar.in,fix_par=0)
    Corr.fix <- Corr.out                    # correlation is fixed
    para.fix <- qle.model1$betahat # intercept is fixed

    QLR.S=sigma2 = array(NA,c(1,m))
    par.out=array(NA,c(ncol(X1)+1,m))
```

```
par.out_c=array(NA,c(ncol(X1)+1,m))
for (k in 1:m){
   R.in <- R
   X.in <- cbind(X1,W[,k])
   par.in <- para.fix
   model1 <- lm(R.in~0+X.in,offset= rep(par.in[1],n))
   par.in <- c(par.in[1],model1$coefficients)
   theta.in <- (X.in%*%par.in[-1])+(par.in[1])
   yvar.in    <- as.vector(var(    R.in-    theta.in))
   qle.model1 <- QLE(beta.in=par.in[-1], X.in=X.in,R.in=R.in,n_i=1,
      Corr=Corr.fix,sigma2=yvar.in,fix_par=par.in[1])
   par.out[,k] <- qle.model1$betahat

   R.in_c <- R
   X.in_c <- cbind(X1_c,W_c[,k])
   model2 <- lm(R.in_c~0+X.in_c,offset= rep(sum(par.in[-2:-6]),n))
   par.in_c <- c(par.in[1],model2$coefficients)
   theta.in_c <- as.vector((X.in_c%*%par.in_c[-1])+(sum(para.fix[-2:-6])))
   yvar.in_c    <- var(R.in_c-theta.in_c)
   qle.model2 <- QLE(beta.in=par.in_c[-1], X.in=X.in_c,R.in=R.in_c,n_i=1,
      Corr=Corr.fix,sigma2=yvar.in_c,fix_par=para.fix[-2:-6])
   par.out_c[,k] <- qle.model2$betahat

   Vinv      <- Corr.fix*yvar.in
   Vinv_c <- Corr.fix*yvar.in_c
   QLR.S[k] <-
   ((2*n)^(-1))*t(theta.in-theta.in_c%*%(Vinv+Vinv_c)%*%(theta.in-theta.in_c)

   U <- solve(t(W[,k])%*%Vinv%*%W[,k],tol=1e-300)%*%W[,k] +
            solve(t(W_c[,k])%*%Vinv_c%*%W_c[,k],tol=1e-300)%*%W_c[,k]
   sigma2_xi[k] <- U %*%Vinv%*%t(U)
   Z_w[k] <- (par.out[ncol(X1)+1,k]-par.out_c[ncol(X1)+1,k])/sqrt(sigma2_xi[k])
   p[k]      <- (1-pnorm(Z_w[k]))*2    }

Max.T2=which.min(QLR.S)
Max.T2.all=c(Max.T2.all,Max.T2)
if(count==20){
   Max.T2=as.numeric( attr( which.max(table(Max.T2.all))[1],"names"));break }
```

```
  if(Max.T2==Max.T) break
  Max.T2 }
Ans.0=Max.T2
Max.T.all=c(Max.T.all,Ans.0)

# Augmented models when at least two jump sets are selected
count=0
repeat{
  count <- count+1; cat(count)
  QLR.S <- array(NA,c(1,m))
  X1      <- cbind(covariate,W[,Max.T.all])
  X1_c <- cbind(covariate,W_c[,Max.T.all])
  if(count!=1){
    X1      <- cbind(covariate,S)
    X1_c <- cbind(covariate,1-S) }
  par.out      <- array(NA,c(ncol(X1)+1,m))
  par.out_c <- array(NA,c(ncol(X1)+1,m))
  QLR.S <- sigma2=sigma2_xi=Z_w=p=array(NA,c(1,m))

  for (k in 1:m){
    if(sum(W[,k])%in%colSums(W[,Max.T.all,drop=F])){next}
    R.in <- R
    X.in <- as.matrix(cbind(X1,W[,k]))
    par.in <- para.fix
    model1 <- lm(R.in~0+X.in,offset= rep(par.in[1],n))
    par.in <- c(par.in[1],model1$coefficients)
    theta.in=(X.in%*%par.in[-1])+(par.in[1])
    yvar.in = as.vector(var(R.in-theta.in))
    qle.model1 <- QLE(beta.in= par.in[-1],X.in=X.in,R.in=R.in,n_i=1,
      Corr=Corr.fix,sigma2=yvar.in,fix_par=par.in[1])
    par.out[,k] <- qle.model1$betahat

    R.in_c <- R
    X.in_c <- as.matrix(cbind(X1_c,W_c[,k]))
    model2 <- lm(R.in_c~0+X.in_c,offset= rep(sum(par.in[-2]),n))
    par.in_c      <- c(par.in[1],model2$coefficients)
    theta.in_c <- as.vector((X.in_c%*%par.in_c[-1])+(sum(par.in[-2])))
    yvar.in_c    <- var(R.in_c-theta.in_c)
```

```
   qle.model2 <- QLE(beta.in= par.in_c[-1],X.in=X.in_c,R.in=R.in_c,n_i=1,
      Corr=Corr.fix,sigma2=yvar.in_c,fix_par= sum(par.in[-2]))
  par.out_c[,k]=qle.model2$betahat
  Vinv=Corr.fix*yvar.in
  Vinv_c=Corr.fix*yvar.in_c
  QLR.S[k] <-
   ((2*n)^(-1))*t(theta.in-theta.in_c)%*%(Vinv+Vinv_c)%*%(theta.in-theta.in_c)
  U <- solve(t(W[,k])%*%Vinv%*%W[,k],tol=1e-300) %*% W[,k] +
        solve(t(W_c[,k])%*%Vinv_c%*%W_c[,k],tol=1e-300) %*% W_c[,k]
  sigma2_xi[k] <- U %*%Vinv%*%t(U)
  Z_w[k] <- (par.out[ncol(X1)+1,k]-par.out_c[ncol(X1)+1,k])/sqrt(sigma2_xi[k])
  p[k] <- (1-pnorm(Z_w[k]))*2 }

if(min(p,na.rm = T)>1-(0.95)^(1/n)) break
Ans.1 <- which.min(QLR.S)
Max.T.all <- c(Max.T.all,Ans.1)
S <- Merg(Max.T.all,W)
R.in <- R
X.in <- as.matrix(cbind(1,covariate,S))
model1 <- lm(R.in~X.in[,-1])
par.in <- model1$coefficients
yvar.in <- as.vector(var(R.in-(X.in%*%par.in)))

cat("Estimated jump sets",Max.T,".")
cov.vgm <- vgm.est(resi.in=(R.in-(X.in%*%par.in)),loc.in=xy,dist.in=L2,
   yvar=yvar.in,cut.off=F)
Corr.out<- cov.vgm
qle.model1 <- QLE(beta.in= par.in,X.in=X.in,R.in=R.in,n_i=1,
   Corr=Corr.out,sigma2=yvar.in,fix_par=0)
Corr.fix=Corr.out
para.fix= qle.model1$betahat }
```