Local Variable Selection and Parameter Estimation of Spatially Varying Coefficient Regression Models

Wesley Brooks

1. Spatially varying coefficients regression

1.1. Model

Consider n data points, observed at sampling locations $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n$, which are distributed in a spatial domain $D \subset \mathbb{R}^2$ according to a density $f(\boldsymbol{s})$. For $i = 1, \dots, n$, let $y(\boldsymbol{s}_i)$ and $\boldsymbol{x}(\boldsymbol{s}_i)$ denote the univariate response variable, and a $(p+1)$-variate vector of covariates measured at location $\boldsymbol{s}_i$, respectively. At each location $\boldsymbol{s}_i$, assume that the outcome is related to the covariates by a linear model where the coefficients $\boldsymbol{\beta}(\boldsymbol{s}_i)$ may be spatially-varying and $\varepsilon(\boldsymbol{s}_i)$ is random error at location $\boldsymbol{s}_i$. That is,

$$y(\boldsymbol{s}_i) = \boldsymbol{x}(\boldsymbol{s}_i)'\boldsymbol{\beta}(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i). \tag{1}$$

Further assume that the error term $\varepsilon(\boldsymbol{s}_i)$ is normally distributed with zero mean and variance σ^2, and that $\varepsilon(\boldsymbol{s}_i)$, $i = 1, \dots, n$ are independent. That is,

$$\varepsilon(\boldsymbol{s}_i) \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right). \tag{2}$$

In the context of nonparametric regression, the boundary-effect bias can be reduced by local polynomial modeling, usually in the form of a locally linear model (Fan and Gijbels, 1996). Here, locally linear coefficients are estimated by augmenting the local design matrix with covariate-by-location interactions in two dimensions as proposed by Wang et al. (2008). The augmented local design matrix at location $\boldsymbol{s}_i$ is

$$\boldsymbol{Z}(\boldsymbol{s}_i) = (\boldsymbol{X} \;\; L_i\boldsymbol{X} \;\; M_i\boldsymbol{X}) \tag{3}$$

where $\boldsymbol{X}$ is the unaugmented matrix of covariates, $L_i = \mathrm{diag}\{s_{i',x} - s_{i,x}\}$ and $M_i = \mathrm{diag}\{s_{i',y} - s_{i,y}\}$ for $i' = 1, \ldots, n$.

1.2. Estimation

The total log-likelihood of the observed data is the sum of the log-likelihood of each individual observation:

$$\ell\left\{\boldsymbol{\beta}\right\} = -(1/2)\sum_{i'=1}^{n}\left[\log\sigma^2 + \sigma^{-2}\left\{y(\boldsymbol{s}_{i'}) - \boldsymbol{z}'(\boldsymbol{s}_{i'})\boldsymbol{\beta}(\boldsymbol{s}_i)\right\}^2\right]. \tag{4}$$

Since there are a total of $n \times 3(p+1)$ free parameters for n observations, the model is not identifiable and it is not possible to directly maximize the total likelihood.

The values of the local coefficients $\boldsymbol{\beta}(\boldsymbol{s})$ are estimated at location $\boldsymbol{s}$ by the weighted likelihood

$$\mathcal{L}\left\{\boldsymbol{\beta}(\boldsymbol{s})\right\} = \prod_{i'=1}^{n}\left\{\left(2\pi\sigma^2\right)^{-1/2}\exp\left[-(1/2)\sigma^{-2}\left\{y(\boldsymbol{s}_{i'}) - \boldsymbol{z}'(\boldsymbol{s}_{i'})\boldsymbol{\beta}(\boldsymbol{s})\right\}^2\right]\right\}^{K_h(\|\boldsymbol{s}-\boldsymbol{s}_{i'}\|)}, \tag{5}$$

where the weights are calculated by a kernel function $K_h(\cdot)$ such as the Epanechnikov kernel:

$$\begin{aligned} K_h(\delta_{ii'}) &= h^{-2}K\left(h^{-1}\delta_{ii'}\right) \\ K(x) &= \begin{cases} (3/4)(1-x^2) & \text{if } \delta_{ii'} < h, \\ 0 & \text{if } \delta_{ii'} \geqslant h. \end{cases} \end{aligned} \tag{6}$$

Thus, the local log-likelihood function is, up to an additive constant:

$$\ell\left\{\boldsymbol{\beta}(\boldsymbol{s})\right\} = -(1/2)\sum_{i'=1}^{n} K_h(\|\boldsymbol{s}-\boldsymbol{s}_{i'}\|)\left[\log\sigma^2 + \sigma^{-2}\left\{y(\boldsymbol{s}_{i'}) - \boldsymbol{z}'(\boldsymbol{s}_{i'})\boldsymbol{\beta}(\boldsymbol{s})\right\}^2\right]. \tag{7}$$

This local likelihood can be maximized by weighted least squares

$$\hat{\boldsymbol{\beta}}(\boldsymbol{s}) = \left\{\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\right\}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y}. \tag{8}$$

From (7), the maximum local likelihood estimate $\hat{\sigma}_i^2$ is:

$$\hat{\sigma}^2(\boldsymbol{s}) = \left\{\sum_{i'=1}^{n} K_h(\|\boldsymbol{s}-\boldsymbol{s}_{i'}\|)\right\}^{-1}\sum_{i'=1}^{n} K_h(\|\boldsymbol{s}-\boldsymbol{s}_{i'}\|)\left\{y(\boldsymbol{s}_{i'}) - \boldsymbol{z}'(\boldsymbol{s}_{i'})\hat{\boldsymbol{\beta}}(\boldsymbol{s})\right\}^2 \tag{9}$$

2. Local variable selection and parameter estimation

2.1. Local variable selection

Local adaptive grouped regularization (LAGR) is explored as a penalty function for local variable selection in SVCR models. The proposed local variable selection with LAGR penalty is an ℓ_1 regularization method for variable selection in regression models (Wang and Leng, 2008; Zou, 2006). The adaptive group lasso selects groups of covariates for inclusion or exclusion in the model. For an SVCR model, each variable group is a covariate and its interactions on location.

2.1.1. *Local variable selection and coefficient estimation with the adaptive group lasso*

The objective function for the LAGR at location $\boldsymbol{s}$ consists of the local log-likelihood and an additive penalty:

$$\begin{aligned}\mathcal{S}\{\boldsymbol{\beta}(\boldsymbol{s})\} &= -2\ell_i\{\boldsymbol{\beta}(\boldsymbol{s})\} + \mathcal{J}_1\{\boldsymbol{\beta}(\boldsymbol{s})\} \\ &= \sum_{i'=1}^{n} K_h(\|\boldsymbol{s}-\boldsymbol{s}_{i'}\|)\left[\log\sigma^2 + \sigma^{-2}\left\{y(\boldsymbol{s}_{i'}) - \boldsymbol{z}'(\boldsymbol{s}_{i'})\boldsymbol{\beta}(\boldsymbol{s}_i)\right\}^2\right] + \lambda_n(\boldsymbol{s})\sum_{j=1}^{p}\|\beta_j(\boldsymbol{s})\|\|\gamma_j(\boldsymbol{s})\|^{-\gamma}\end{aligned} \tag{10}$$

where $\sum_{i'=1}^{n} K_h(\|\boldsymbol{s}-\boldsymbol{s}_i\|)\left\{y(\boldsymbol{s}_{i'}) - \boldsymbol{z}'(\boldsymbol{s}_{i'})\boldsymbol{\beta}(\boldsymbol{s})\right\}^2$ is the weighted sum of squares minimized by traditional GWR, and $\mathcal{J}_1\{\boldsymbol{\beta}(\boldsymbol{s})\} = \lambda_n(\boldsymbol{s})\sum_{j=1}^{p}\|\beta_j(\boldsymbol{s})\|\|\tilde{\beta}_j(\boldsymbol{s})\|^{-\gamma}$ is the LAGR penalty. With the vector of unpenalized local coefficients $\tilde{\boldsymbol{\beta}}(\boldsymbol{s})$, the LAGR penalty for the jth group of coefficients $\beta_j(\boldsymbol{s})$ at location $\boldsymbol{s}$ is $\lambda_n(\boldsymbol{s})\|\tilde{\beta}_j(\boldsymbol{s})\|^{-\gamma}$, where $\lambda_n(\boldsymbol{s}) > 0$ is a the local tuning parameter applied to all coefficients at location $\boldsymbol{s}$ and $\boldsymbol{\gamma}(\boldsymbol{s}) = \{\gamma_1(\boldsymbol{s}), \ldots, \gamma_p(\boldsymbol{s})\}'$ is the vector of adaptive weights at location $\boldsymbol{s}$.

3. Asymptotic properties

Consider a local model at location $\boldsymbol{s}$ where there are $p_0 < p$ covariates with nonzero local regression coefficients. Without loss of generality, assume these are covariates $1, \ldots, p_0$.

Let $a_n = \max\{\lambda_n(\boldsymbol{s})\|\tilde{\beta}_j(\boldsymbol{s})\|^{-\gamma}, j \leqslant p_0\}$ be the largest penalty applied to a covariate group whose true coefficient norm is nonzero, and $b_n = \min\{\lambda_n(\boldsymbol{s})\|\tilde{\beta}_j(\boldsymbol{s})\|^{-\gamma}, j > p_0\}$ be the smallest penalty applied to a covariate group whose true coefficient norm is zero.

3.1. *Asymptotic normality*

Theorem 3.1. *If* $h^{-1}n^{-1/2}a_n \xrightarrow{p} 0$ *and* $hn^{-1/2}b_n \xrightarrow{p} \infty$ *then* $\hat{\boldsymbol{\beta}}(\boldsymbol{s}) - \boldsymbol{\beta}(\boldsymbol{s}) - \frac{\kappa_2 h^2}{2\kappa_0}\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{vv}(\boldsymbol{s})\} = O_p(n^{-1/2}h^{-1})$

Proof. The idea of the proof is to show that the objective being minimized achieves a unique minimum, which must be $\hat{\boldsymbol{\beta}}(\boldsymbol{s})$.

The order of convergence is $hn^{1/2}$ where $h = O(n^{-1/6})$.

We find the limiting distribution of the estimator:

$$\begin{aligned}
V_4^{(n)}(\boldsymbol{u}) &= Q\left\{\boldsymbol{\beta}(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}\right\} - Q\left\{\boldsymbol{\beta}(\boldsymbol{s})\right\} \\
&= (1/2)\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\left\{\boldsymbol{\beta}(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}\right\}\right]^T \boldsymbol{W}(\boldsymbol{s})\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\left\{\boldsymbol{\beta}(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}\right\}\right] \\
&\quad + \sum_{j=1}^{p} \lambda_j \|\boldsymbol{\beta}(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}_j\| \\
&\quad - (1/2)\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}^T \boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\} - \sum_{j=1}^{p} \lambda_j \|\boldsymbol{\beta}(\boldsymbol{s})\| \\
&= (1/2)\boldsymbol{u}^T\left\{h^{-2}n^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\right\}\boldsymbol{u} - \boldsymbol{u}^T\left[h^{-1}n^{-1/2}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right] \\
&\quad + \sum_{j=1}^{p} n^{-1/2}\lambda_j n^{1/2}\left\{\|\boldsymbol{\beta}_j(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}_j\| - \|\boldsymbol{\beta}_j(\boldsymbol{s})\|\right\} \qquad (11)
\end{aligned}$$

Note the different limiting behavior of the third term between the cases $j \leqslant p_0$ and $j > p_0$:

Case $j \leqslant p_0$. If $j \leqslant p_0$ then $n^{-1/2}\lambda_j \to n^{-1/2}\lambda\|\boldsymbol{\beta}_j(\boldsymbol{s})\|^{-\gamma}$ and $|\sqrt{n}\left\{\|\boldsymbol{\beta}_j(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}_j\| - \|\boldsymbol{\beta}_j(\boldsymbol{s})\|\right\}| \leqslant h^{-1}\|\boldsymbol{u}_j\|$ so $\lim_{n\to\infty} \lambda_j\left(\|\boldsymbol{\beta}_j(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}_j\| - \|\boldsymbol{\beta}_j(\boldsymbol{s})\|\right) \leqslant h^{-1}n^{-1/2}\lambda_j\|\boldsymbol{u}_j\| \leqslant h^{-1}n^{-1/2}a_n\|\boldsymbol{u}_j\| \to 0$

Case $j > p_0$. If $j > p_0$ then $\lambda_j\left(\|\boldsymbol{\beta}_j(\boldsymbol{s}) + h^{-1}n^{-1/2}\boldsymbol{u}_j\| - \|\boldsymbol{\beta}_j(\boldsymbol{s})\|\right) = \lambda_j h^{-1}n^{-1/2}\|\boldsymbol{u}_j\|$.

And note that $h = O(n^{-1/6})$ so that if $hn^{-1/2}b_n \xrightarrow{p} \infty$ then $h^{-1}n^{-1/2}b_n \xrightarrow{p} \infty$.

Now, if $\|\boldsymbol{u}_j\| \neq 0$ then $h^{-1}n^{-1/2}\lambda_j\|\boldsymbol{u}_j\| \geqslant h^{-1}n^{-1/2}b_n\|\boldsymbol{u}_j\| \to \infty$. On the other hand, if $\|\boldsymbol{u}_j\| = 0$ then $h^{-1}n^{-1/2}\lambda_j\|\boldsymbol{u}_j\| = 0$.

Thus, the limit of $V_4^{(n)}(\boldsymbol{u})$ is the same as the limit of $V_4^{*(n)}(\boldsymbol{u})$ where

$$V_4^{*(n)}(\boldsymbol{u}) = \begin{cases} (1/2)\boldsymbol{u}^T\left\{h^{-2}n^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\right\}\boldsymbol{u} - \boldsymbol{u}^T\left[h^{-1}n^{-1/2}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right] & \text{if } \|\boldsymbol{u}_j\| = 0 \ \forall j > p_0 \\ \infty & \text{otherwise} \end{cases}$$

And $V_4^{*(n)}(\boldsymbol{u})$ is convex and is minimized at $\hat{\boldsymbol{u}}^{(n)}$:

$$
\begin{aligned}
0 &= \left\{h^{-2}n^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\right\}\hat{\boldsymbol{u}}^{(n)} - \left[h^{-1}n^{-1/2}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right] \\
\therefore \hat{\boldsymbol{u}}^{(n)} &= \left\{n^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\right\}^{-1}\left[hn^{-1/2}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right]
\end{aligned} \tag{12}
$$

By the epiconvergence results of Geyer (1994) and Knight and Fu (2000), the minimizer of the limiting function is the limit of the minimizers $\hat{\boldsymbol{u}}^{(n)}$. And since, by Lemma 2 of Sun et al. (2014),

$$
\hat{\boldsymbol{u}}^{(n)} \xrightarrow{d} N\left(0, f(\boldsymbol{s})\kappa_0^{-2}\nu_0\sigma^2\Psi^{-1}\right) \tag{13}
$$

the result is proven. □

3.2. Selection

Theorem 3.2. *If $hn^{-1/2}b_n \xrightarrow{p} \infty$ then $P\left\{\hat{\boldsymbol{\beta}}_{(b)}(\boldsymbol{s}) = 0\right\} \to 1$.*

Proof. We showed in Theorem 3.1 that the

The proof is by contradiction.

Recall that the objective to be minimized by $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})$ is

$$
Q\left\{\hat{\boldsymbol{\beta}}(\boldsymbol{s})\right\} = (1/2)\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\hat{\boldsymbol{\beta}}(\boldsymbol{s})\right\}^T \boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\hat{\boldsymbol{\beta}}(\boldsymbol{s})\right\} + \sum_{j=1}^{p}\lambda_j\|\hat{\boldsymbol{\beta}}_p(\boldsymbol{s})\| \tag{14}
$$

Let $\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}) \neq 0$. Then (14) is differentiable w.r.t. $\boldsymbol{\beta}_p(\boldsymbol{s})$ and Q is maximized at

$$
\begin{aligned}
0 &= \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}_{(-p)}(\boldsymbol{s})\hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s}) - \boldsymbol{Z}_{(p)}(\boldsymbol{s})\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\right\} - \lambda_p\frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|} \\
&= \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{vv}(\boldsymbol{s})\right\}\right] \\
&\qquad + \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(-p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(-p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s})\right] \\
&\qquad + \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(p)}(\boldsymbol{s})\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\right] \\
&\qquad - \lambda_p\frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|}
\end{aligned} \tag{15}
$$

So

$$
\begin{aligned}
\frac{h}{\sqrt{n}}\lambda_p \frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|} &= \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\frac{h}{\sqrt{n}}\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{vv}(\boldsymbol{s})\right\}\right] \\
&+ \left\{n^{-1}\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\right\} h\sqrt{n}\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(-p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(-p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s})\right] \\
&+ \left\{n^{-1}\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\right\} h\sqrt{n}\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\right]
\end{aligned}
\tag{16}
$$

From Lemma 2 of Sun et al. (2014), $\left\{n^{-1}\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\right\} = O_p(1)$ and $\left\{n^{-1}\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(p)}(\boldsymbol{s})\right\} = O_p(1)$.

From Theorem 3 of Sun et al. (2014), we have that $h\sqrt{n}\left[\hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s}) - \boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\beta_{(p),uu}(\boldsymbol{s}) + \beta_{(p),vv}(\boldsymbol{s})\right\}\right] = O_p(1)$ and $h\sqrt{n}\left[\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}) - \boldsymbol{\beta}_{(p)}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\beta_{(p),uu}(\boldsymbol{s}) + \beta_{(p),vv}(\boldsymbol{s})\right\}\right] = O_p(1)$.

So the second and third terms of the sum in (16) are $O_p(1)$.

We showed in the proof of 3.1 that $h\sqrt{n}\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{vv}(\boldsymbol{s})\right\}\right] = O_p(1)$.

The three terms of the sum to the right of the equals sign in (16) are $O_p(1)$, so for $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})$ to be a solution, we must have that $hn^{-1/2}\lambda_p\frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|} = O_p(1)$.

But since by assumption $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i) \neq 0$, there must be some $k \in \{1, \ldots, d_p\}$ such that $|\hat{\beta}_{(p),k}(\boldsymbol{s})| = \max\{|\hat{\beta}_{(p),k'}(\boldsymbol{s})| : 1 \leqslant k' \leqslant d_p\}$. And for this k, we have that $|\hat{\beta}_{(p),k}(\boldsymbol{s})|/\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\| \geqslant 1/\sqrt{d_p} > 0$.

Now since $hn^{-1/2}b_n \to \infty$, we have that $hn^{-1/2}\lambda_p\frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|} \geqslant hn^{-1/2}b_n d_p^{-1/2} \to \infty$ and therefore the term to the left of the equals sign dominates the sum to the right of the equals sign in (16). So for large enough n, $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}) \neq 0$ cannot maximize Q.

So $P\left\{\hat{\boldsymbol{\beta}}_{(b)}(\boldsymbol{s}) = 0\right\} \to 1$. □

3.3. A note on rates

To prove the oracle properties of LAGR, we assumed that $h^{-1}n^{-1/2}a_n \xrightarrow{p} 0$ and $hn^{-1/2}b_n \xrightarrow{p} \infty$.

Therefore, $h^{-1}n^{-1/2}\lambda_n \to 0$ for $j \leqslant p_0$ and $hn^{-1/2}\lambda_n\|\boldsymbol{\beta}_j(\boldsymbol{s})\|^{-\gamma} \to \infty$ for $j > p_0$.

We require a λ_n that can satisfy both assumptions. Suppose $\lambda_n = n^{\alpha}$, and recall that $h = O(n^{-1/6})$.

Then $h^{-1}n^{-1/2}\lambda_n = O(n^{-1/3+\alpha})$ and $hn^{-1/2}\lambda_n(h\sqrt{n})^{\gamma} = O(n^{-2/3+\alpha+\gamma/3})$.

So $(2-\gamma)/3 < \alpha < 1/3$, which can only be satisfied for $\gamma > 1$.

4. References

References

Fan, J. and I. Gijbels (1996). *Local Polynomial Modeling and its Applications*. Chapman and Hall, London.

Geyer, C. J. (1994). On the asymptotics of constrained m-estimation. *Annals of Statistics 22*, 1993–2010.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *Annals of Statistics 28*, 1356–1378.

Sun, Y., H. Yan, W. Zhang, and Z. Lu (2014). A semiparametric spatial dynamic model.

Wang, H. and C. Leng (2008). A note on adaptive group lasso. *Computational Statistics and Data Analysis 52*, 5277–5286.

Wang, N., C.-L. Mei, and X.-D. Yan (2008). Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. *Environment and Planning A 40*, 986–1005.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*, 1418–1429.