

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

Wesley Brooks

0.1. Model

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

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

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

$$\varepsilon(\mathbf{s}_i) \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2). \quad (2)$$

Thus, conditional on the design matrix \mathbf{X} , observations of the response variable at different locations are independent of each other.

An SVCR model that estimates the regression coefficients as locally constant, as in the class of Nadaraya-Watson kernel smoothers (?), suffers the problem of biased estimation that is common

to that class of models - particularly where there is a gradient to the coefficient surface at the boundary of the domain (?).

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 (?). Here, locally linear coefficients are estimated by augmenting the local design matrix with covariate-by-location interactions in two dimensions as proposed by ?. The augmented local design matrix at location \mathbf{s}_i is

$$\mathbf{Z}(\mathbf{s}_i) = (\mathbf{X} \quad L(\mathbf{s}_i) \mathbf{X} \quad M(\mathbf{s}_i) \mathbf{X}) \quad (3)$$

where \mathbf{X} is the unaugmented matrix of covariates, $L(\mathbf{s}_i) = \text{diag}\{(\mathbf{s}_{i'} - \mathbf{s}_i)_1\}$ and $M(\mathbf{s}_i) = \text{diag}\{(\mathbf{s}_{i'} - \mathbf{s}_i)_2\}$ for $i' = 1, \dots, n$.

0.2. Estimation

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

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

Since there are a total of $n \times 3(p+1)$ 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 $\boldsymbol{\beta}(\mathbf{s}_i)$ are estimated by the weighted likelihood

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

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

$$w_{ii'} = K_h(\delta_{ii'}) = h^{-2} K(h^{-1} \delta_{ii'})$$

$$K(x) = \begin{cases} (3/4)(1 - x^2) & \text{if } \delta_{ii'} < h, \\ 0 & \text{if } \delta_{ii'} \geq h. \end{cases} \quad (6)$$

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

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

This local likelihood can be maximized by weighted least squares

$$\hat{\boldsymbol{\beta}}(\mathbf{s}_i) = \{\mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \mathbf{Z}(\mathbf{s}_i)\}^{-1} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \mathbf{Y}. \quad (8)$$

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

$$\hat{\sigma}^2(\mathbf{s}_i) = \left(\sum_{i'=1}^n w_{ii'} \right)^{-1} \sum_{i'=1}^n w_{ii'} \{y(\mathbf{s}_{i'}) - \mathbf{z}'(\mathbf{s}_{i'}) \hat{\boldsymbol{\beta}}(\mathbf{s}_i)\}^2 \quad (9)$$

1. Asymptotics

1.1. Consistency

Theorem 1.1. *If $\sqrt{n} a_n \xrightarrow{p} 0$ then $\hat{\boldsymbol{\beta}}(\mathbf{s}_i) - \boldsymbol{\beta}(\mathbf{s}_i) - \frac{\kappa_2 h^2}{2\kappa_0} \{\boldsymbol{\beta}_{uu}(\mathbf{s}_i) + \boldsymbol{\beta}_{vv}(\mathbf{s}_i)\} = O_p(n^{-1/2} h^{-1})$*

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

The order of convergence is $n^{1/2} h$ where $h = O(n^{-1/6})$ so that the rate of convergence is $n^{1/3}$.

To show: that for any ϵ , there is a sufficiently large constant C such that

$$\liminf_n P \left[\inf_{u \in \mathcal{R}: \|u\|=C} Q \{ \boldsymbol{\beta}(\mathbf{s}_i) + n^{-1/2} \mathbf{u} \} > Q \{ \boldsymbol{\beta}(\mathbf{s}_i) \} \right] > 1 - \epsilon$$

We show the result:

$$\begin{aligned}
Q(\beta(\mathbf{s}_i) + n^{-1/2}\mathbf{u}) - Q(\beta(\mathbf{s}_i)) &= (1/2) \left[\mathbf{Y} - \mathbf{Z}(\mathbf{s}_i) \left\{ \beta(\mathbf{s}_i) + n^{-1/2}\mathbf{u} \right\} \right]^T \mathbf{W}(\mathbf{s}_i) \left[\mathbf{Y} - \mathbf{Z}(\mathbf{s}_i) \left\{ \beta(\mathbf{s}_i) + n^{-1/2}\mathbf{u} \right\} \right] \\
&\quad + n \sum_{j=1}^p \lambda_j \|\beta(\mathbf{s}_i) + n^{-1/2}\mathbf{u}\| \\
&\quad - (1/2) \left\{ \mathbf{Y} - \mathbf{Z}\beta(\mathbf{s}_i) \right\}^T \mathbf{W}(\mathbf{s}_i) \left\{ \mathbf{Y} - \mathbf{Z}\beta(\mathbf{s}_i) \right\} + n \sum_{j=1}^p \lambda_j \|\beta(\mathbf{s}_i)\| \\
&= (1/2) \mathbf{u}^T \left\{ \frac{1}{n} \mathbf{Z}^T \mathbf{W}(\mathbf{s}_i) \mathbf{Z}(\mathbf{s}_i) \right\} \mathbf{u} - \mathbf{u}^T \left[n^{-1/2} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \left\{ \mathbf{Y} - \mathbf{Z}(\mathbf{s}_i) \beta(\mathbf{s}_i) \right\} \right] \\
&\quad + n \sum_{j=1}^p \lambda_j \|\beta_j(\mathbf{s}_i) + n^{-1/2}\mathbf{u}\| - n \sum_{j=1}^p \lambda_j \|\beta_j(\mathbf{s}_i)\| \\
&= (1/2) \mathbf{u}^T \left\{ \frac{1}{n} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \mathbf{Z}(\mathbf{s}_i) \right\} \mathbf{u} - \mathbf{u}^T \left[n^{-1/2} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \left\{ \mathbf{Y} - \mathbf{Z}(\mathbf{s}_i) \beta(\mathbf{s}_i) \right\} \right] \\
&\quad + n \sum_{j=1}^p \lambda_j \|\beta_j(\mathbf{s}_i) + n^{-1/2}\mathbf{u}\| - n \sum_{j=1}^{p_0} \lambda_j \|\beta_j(\mathbf{s}_i)\| \\
&\geq (1/2) \mathbf{u}^T \left\{ \frac{1}{n} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \mathbf{Z}(\mathbf{s}_i) \right\} \mathbf{u} - \mathbf{u}^T \left[n^{-1/2} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \left\{ \mathbf{Y} - \mathbf{Z}(\mathbf{s}_i) \beta(\mathbf{s}_i) \right\} \right] \\
&\quad + n \sum_{j=1}^{p_0} \lambda_j (\|\beta_j(\mathbf{s}_i) + n^{-1/2}\mathbf{u}\| - \|\beta_j(\mathbf{s}_i)\|) \\
&\geq (1/2) \mathbf{u}^T \left\{ \frac{1}{n} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \mathbf{Z}(\mathbf{s}_i) \right\} \mathbf{u} - \mathbf{u}^T \left[n^{-1/2} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \left\{ \mathbf{Y} - \mathbf{Z}(\mathbf{s}_i) \beta(\mathbf{s}_i) \right\} \right] \\
&\quad - p_0 \sqrt{n} a_n \|\mathbf{u}\|
\end{aligned} \tag{10}$$

□

We'll consider the terms of the sum in (10) separately.

First term.. By Lemma 2 of ?, $\frac{1}{n} \mathbf{Z}^T(\mathbf{s}_i) \mathbf{W}(\mathbf{s}_i) \mathbf{Z}(\mathbf{s}_i) \xrightarrow{p} \Omega$, so the first term in (10) converges to $\mathbf{u}^T \Omega \mathbf{u}$, a quadratic form in \mathbf{u} .

Second term.. By a first-order Taylor expansion, we have that $\beta(\mathbf{s}_i) = \beta(\mathbf{s}) + \nabla\beta(\boldsymbol{\xi}_i)(\mathbf{s}_i - \mathbf{s})$ where $\boldsymbol{\xi}_i = \mathbf{s} + \theta(\mathbf{s}_i - \mathbf{s})$ and $\theta \in [0, 1]$ for $i = 1, \dots, n$. So

$$\mathbf{Y} - \mathbf{Z}(\mathbf{s}_i)\beta(\mathbf{s}_i) = \mathbf{m} + \boldsymbol{\varepsilon} - \mathbf{Z}(\mathbf{s}_i)\beta(\mathbf{s}_i)$$

and so the second term of (10) is

$$\mathbf{u}^T \left[n^{-1/2} \mathbf{Z}^T(\mathbf{s}) \mathbf{W}(\mathbf{s}) \{ \mathbf{m} + \boldsymbol{\varepsilon} - \mathbf{Z}(\mathbf{s})\beta(\mathbf{s}) \} \right]. \quad (11)$$

We wish to show that (11) is $O_p(1)$. Let $\{\mathbf{A}\}_j$ be the j th column of the matrix \mathbf{A} as a column vector, and $\{\mathbf{A}\}_k^T$ be the k th row of the matrix \mathbf{A} as a row vector. Now, taking the three terms of the sum separately:

First term. Find the expectation and variance of the i th term in the sum $n^{-1/2} \mathbf{H}^{-1} \mathbf{Z}^T(\mathbf{s}) \mathbf{W}(\mathbf{s}) \mathbf{m}$:

$$\begin{aligned} & \mathbf{H}^{-1} \{ \mathbf{Z}^T(\mathbf{s}) \}_i \{ \mathbf{W}(\mathbf{s}) \}_{ii} m(\mathbf{s}_i) \\ &= K_h(\|\mathbf{s} - \mathbf{s}_i\|) \mathbf{H}^{-1} \{ \mathbf{Z}^T(\mathbf{s}) \}_i \{ \mathbf{Z}(\mathbf{s}_i) \}_i^T \boldsymbol{\gamma}(\mathbf{s}_i) \\ &= K_h(\|\mathbf{s} - \mathbf{s}_i\|) \begin{pmatrix} X_1^2(\mathbf{s}_i) & \dots & X_1(\mathbf{s}_i)X_p(\mathbf{s}_i) & \mathbf{0}_{1 \times 2p} \\ \vdots & \ddots & \vdots & \vdots \\ X_1(\mathbf{s}_i)X_p(\mathbf{s}_i) & \dots & X_p^2(\mathbf{s}_i) & \mathbf{0}_{1 \times 2p} \\ \mathbf{0}_{2p \times 1} & \dots & \mathbf{0}_{2p \times 1} & \mathbf{0}_{2p \times 2p} \end{pmatrix} \boldsymbol{\gamma}(\mathbf{s}_i) \\ &= K_h(\|\mathbf{s} - \mathbf{s}_i\|) \begin{pmatrix} \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i) & \mathbf{0}_{p \times 2p} \\ \mathbf{0}_{2p \times p} & \mathbf{0}_{2p \times 2p} \end{pmatrix} \boldsymbol{\gamma}(\mathbf{s}_i) \end{aligned} \quad (12)$$

So the expectation of (12) is:

$$\begin{aligned}
& \mathbb{E}[\mathbf{H}^{-1} \{\mathbf{Z}^T(\mathbf{s})\}_i \{\mathbf{W}(\mathbf{s})\}_{ii} \{\mathbf{Z}(\mathbf{s}_i)\}_i^T \gamma(\mathbf{s}_i)] \\
&= \mathbb{E} \left(\begin{pmatrix} \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i) & \mathbf{0}_{p \times 2p} \\ \mathbf{0}_{2p \times p} & \mathbf{0}_{2p \times 2p} \end{pmatrix} \int K_h(\|\mathbf{s} - \mathbf{t}\|) \gamma(\mathbf{t}) f(\mathbf{t}) d\mathbf{t} \right. \\
&= \left. \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{0}_{p \times 2p} \\ \mathbf{0}_{2p \times p} & \mathbf{0}_{2p \times 2p} \end{pmatrix} \gamma(\mathbf{s}) f(\mathbf{s}) \right) \tag{13}
\end{aligned}$$

And the variance of (12) is:

$$\begin{aligned}
& \mathbb{E}[\mathbf{H}^{-1} \{\mathbf{Z}^T(\mathbf{s})\}_i \{\mathbf{W}(\mathbf{s})\}_{ii} \{\mathbf{Z}(\mathbf{s}_i)\}_i^T \gamma(\mathbf{s}_i) \gamma^T(\mathbf{s}_i) \{\mathbf{Z}^T(\mathbf{s}_i)\}_i \{\mathbf{W}(\mathbf{s})\}_{ii} \{\mathbf{Z}(\mathbf{s})\}_i^T \mathbf{H}^{-1}] \\
&= \mathbb{E} \left(\begin{pmatrix} \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i) & \mathbf{0}_{p \times 2p} \\ \mathbf{0}_{2p \times p} & \mathbf{0}_{2p \times 2p} \end{pmatrix} \int K_h(\|\mathbf{s} - \mathbf{t}\|) \gamma(\mathbf{t}) f(\mathbf{t}) d\mathbf{t} \right. \\
&= \left. \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{0}_{p \times 2p} \\ \mathbf{0}_{2p \times p} & \mathbf{0}_{2p \times 2p} \end{pmatrix} \gamma(\mathbf{s}) f(\mathbf{s}) \right) \tag{14}
\end{aligned}$$

Second term. Find the expectation and variance of the i th term in the sum $n^{-1/2} \mathbf{H}^{-1} \mathbf{Z}^T(\mathbf{s}) \mathbf{W}(\mathbf{s}) \mathbf{Z}(\mathbf{s}) \gamma(\mathbf{s})$:

Third term. Find the expectation and variance of the i th term in the sum $n^{-1/2} \mathbf{H}^{-1} \mathbf{Z}^T(\mathbf{s}) \mathbf{W}(\mathbf{s}) \boldsymbol{\varepsilon}$:

$$\begin{aligned}
& \mathbf{H}^{-1} \{ \mathbf{Z}^T(\mathbf{s}) \}_i \{ \mathbf{W}(\mathbf{s}) \}_{ii} \varepsilon(\mathbf{s}_i) \\
&= K_h(\|\mathbf{s} - \mathbf{s}_i\|) \mathbf{H}^{-1} \{ \mathbf{Z}^T(\mathbf{s}) \}_i \varepsilon(\mathbf{s}_i) \\
&= K_h(\|\mathbf{s} - \mathbf{s}_i\|) \begin{pmatrix} X_1(\mathbf{s}_i) \\ \vdots \\ X_p(\mathbf{s}_i) \\ X_1(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1 \\ \vdots \\ X_p(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1 \\ X_1(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2 \\ \vdots \\ X_p(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2 \end{pmatrix} \varepsilon(\mathbf{s}_i) \\
&= K_h(\|\mathbf{s} - \mathbf{s}_i\|) \begin{pmatrix} \mathbf{X}(\mathbf{s}_i) \\ \mathbf{X}(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1 \\ \mathbf{X}(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2 \end{pmatrix} \varepsilon(\mathbf{s}_i) \tag{15}
\end{aligned}$$

So the expectation of (??) is:

$$\begin{aligned}
& \mathbb{E}[\mathbf{H}^{-1} \{\mathbf{Z}^T(\mathbf{s})\}_i \{\mathbf{W}(\mathbf{s})\}_{ii} \varepsilon(\mathbf{s}_i)] \\
&= \mathbb{E} \begin{pmatrix} \mathbf{X}(\mathbf{s}_i) \\ \mathbf{X}(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1 \\ \mathbf{X}(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2 \end{pmatrix} \mathbb{E} \varepsilon(\mathbf{s}_i) \int K_h(\|\mathbf{s} - \mathbf{t}\|) f(\mathbf{t}) \partial \mathbf{t} \\
&= \begin{pmatrix} \boldsymbol{\mu}(\mathbf{s}_i) \\ \boldsymbol{\mu}(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1 \\ \boldsymbol{\mu}(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2 \end{pmatrix} \times 0 \times f(\mathbf{s}) \\
&= 0
\end{aligned} \tag{16}$$

And the variance of (??) is:

$$\begin{aligned}
& \mathbb{E}[\mathbf{H}^{-1} \{\mathbf{Z}^T(\mathbf{s})\}_i \{\mathbf{W}(\mathbf{s})\}_{ii} \varepsilon^2(\mathbf{s}_i) \{\mathbf{W}(\mathbf{s})\}_{ii} \{\mathbf{Z}(\mathbf{s})\}_i^T \mathbf{H}^{-1}] \\
&= \mathbb{E} \begin{pmatrix} \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i) & \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1 & \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2 \\ h^{-1} \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1 & \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1^2 & \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1(\mathbf{s}_i - \mathbf{s})_2 \\ h^{-1} \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2 & \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_1(\mathbf{s}_i - \mathbf{s})_2 & \mathbf{X}(\mathbf{s}_i) \mathbf{X}^T(\mathbf{s}_i)(\mathbf{s}_i - \mathbf{s})_2^2 \end{pmatrix} \\
&\quad \times \mathbb{E} \varepsilon^2(\mathbf{s}_i) \int K_h^2(\|\mathbf{s} - \mathbf{t}\|) f(\mathbf{t}) \partial \mathbf{t} \\
&= \begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_1 & \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_2 \\ h^{-1} \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_1 & \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_1^2 & \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_1(\mathbf{s}_i - \mathbf{s})_2 \\ h^{-1} \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_2 & \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_1(\mathbf{s}_i - \mathbf{s})_2 & \boldsymbol{\Psi}(\mathbf{s}_i - \mathbf{s})_2^2 \end{pmatrix} \times \sigma^2 \times f(\mathbf{s}) \\
&= 0
\end{aligned} \tag{17}$$

Third term.. By assumption, $p_0 \sqrt{n} a_n = O(\sqrt{n} a_n) = o_p(1)$.

So the quadratic term dominates the sum, implying that the difference $Q\{\beta(s_i) + n^{-1/2}\mathbf{u}\} > Q\{\beta(s_i)\}$ is positive, which proves the result.

1.2. Selection

Theorem 1.2. *If $\sqrt{na_n} \xrightarrow{p} 0$ and $\sqrt{nb_n} \xrightarrow{p} \infty$ then $P\{\hat{\beta}_{(b)}(s_i) = 0\} \rightarrow 1$.*

Proof. The proof is by contradiction. Specifically, we show that if the statement of the theorem does not hold, then the MLE $\hat{\beta}(s_i)$ cannot be a maximum of the likelihood.

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

$$Q\{\beta(s_i)\} = (1/2)\{\mathbf{Y} - \mathbf{Z}(s_i)\beta(s_i)\}^T \mathbf{W}(s_i)\{\mathbf{Y} - \mathbf{Z}(s_i)\beta(s_i)\} + n \sum_{j=1}^p \lambda_j \|\beta_p(s_i)\| \quad (18)$$

Let $\hat{\beta}_p(s_i) \neq 0$. Then (??) is differentiable w.r.t. $\beta_p(s_i)$ and Q is maximized at

$$\begin{aligned} 0 &= \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \left\{ \mathbf{Y} - \mathbf{X}_{(-p)} \hat{\beta}_{-p}(s_i) - \mathbf{X}_{(p)} \hat{\beta}_p(s_i) \right\} + n \lambda_p \frac{\hat{\beta}_p(s_i)}{\|\beta_p(s_i)\|} \\ &= \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \left\{ \mathbf{Y} - \mathbf{X} \beta(s_i) + \mathbf{X} \beta(s_i) - \mathbf{X}_{(-p)} \hat{\beta}_{-p}(s_i) - \mathbf{X}_{(p)} \hat{\beta}_p(s_i) \right\} + n \lambda_p \frac{\hat{\beta}_p(s_i)}{\|\beta_p(s_i)\|} \\ &= \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \{\mathbf{Y} - \mathbf{X} \beta(s_i)\} + \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \left\{ \mathbf{X}_{(-p)} \beta_{-p}(s_i) - \mathbf{X}_{(-p)} \hat{\beta}_{-p}(s_i) \right\} \\ &\quad + \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \left\{ \mathbf{X}_{(-p)} \beta_{-p}(s_i) - \mathbf{X}_{(p)} \hat{\beta}_p(s_i) \right\} + n \lambda_p \frac{\hat{\beta}_p(s_i)}{\|\beta_p(s_i)\|} \\ &= \sqrt{f(s_i) h^2 n^{-1}} \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \{\mathbf{Y} - \mathbf{X} \beta(s_i)\} + \sqrt{f(s_i) h^2 n^{-1}} \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \mathbf{X}_{(-p)} \left\{ \beta_{-p}(s_i) - \hat{\beta}_{-p}(s_i) \right\} \\ &\quad + \sqrt{f(s_i) h^2 n^{-1}} \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \mathbf{X}_{(-p)} \left\{ \beta_{-p}(s_i) - \hat{\beta}_p(s_i) \right\} + \sqrt{f(s_i) h^2 n} \lambda_p \frac{\hat{\beta}_p(s_i)}{\|\beta_p(s_i)\|} \\ &= \sqrt{f(s_i) h^2 n^{-1}} \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \{\mathbf{Y} - \mathbf{X} \beta(s_i)\} + n^{-1} \left\{ \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \mathbf{X}_{(-p)} \right\} \sqrt{f(s_i) h^2 n} \left\{ \beta_{-p}(s_i) - \hat{\beta}_{-p}(s_i) \right\} \\ &\quad + n^{-1} \left\{ \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \mathbf{X}_{(-p)} \right\} \sqrt{f(s_i) h^2 n} \left\{ \beta_{-p}(s_i) - \hat{\beta}_p(s_i) \right\} + \sqrt{f(s_i) h^2 n} \lambda_p \frac{\hat{\beta}_p(s_i)}{\|\beta_p(s_i)\|} \end{aligned} \quad (19)$$

From Lemma 2 of ?, $n^{-1} \left\{ \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \mathbf{X}_{(-p)} \right\} = O_p(1)$ and $n^{-1} \left\{ \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \mathbf{X}_{(p)} \right\} = O_p(1)$. From Theorem 3 of ?, we have that $\sqrt{f(s_i) h^2 n} \left\{ \beta_{(-p)}(s_i) - \hat{\beta}_{(-p)}(s_i) \right\} = O_p(1)$ and $\sqrt{f(s_i) h^2 n} \left\{ \beta_{(p)}(s_i) - \hat{\beta}_{(p)}(s_i) \right\} = O_p(1)$. So the second and third terms of the sum in (??) are $O_p(1)$. We showed in the proof of 1.1 that $\sqrt{f(s_i) h^2 n^{-1}} \mathbf{X}_{(p)}^T \mathbf{W}(s_i) \{\mathbf{Y} - \mathbf{X} \beta(s_i)\} = O_p(1)$.

Because the first three terms of the sum in ?? are $O_p(1)$, for $\hat{\beta}_{(p)}(s_i)$ to be a solution, we must have that $\sqrt{f(s_i) h^2 n} \lambda_p \frac{\hat{\beta}_{(p)}(s_i)}{\|\beta_{(p)}(s_i)\|} = O_p(1)$.

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

Now since $\sqrt{nb_n} \rightarrow \infty$, we have that $\sqrt{f(\mathbf{s}_i)h^2n}\lambda_p \frac{\hat{\beta}_{(p)}(\mathbf{s}_i)}{\|\hat{\beta}_{(p)}(\mathbf{s}_i)\|}$ is unbounded and therefore dominates the $O_p(1)$ terms of the sum in (??). So for large enough n , $\hat{\beta}_{(p)}(\mathbf{s}_i) \neq 0$ cannot maximize Q . \square

1.3. Oracle property

Here we show that the estimation accuracy is just as good as if the relevant predictor groups were specified in advance.

Theorem 1.3. *If $\sqrt{na_n} \rightarrow 0$ and $\sqrt{nb_n} \rightarrow \infty$, then $\sqrt{nh^2f(\mathbf{s})} \left(\hat{\beta}_{(a)}(\mathbf{s}_i) - \beta_{(a)}(\mathbf{s}_i) - \frac{\kappa_2 h^2}{2\kappa_0} \{\beta_{uu}(\mathbf{s}_i) + \beta_{vv}(\mathbf{s}_i)\} \right) \xrightarrow{d} N(0, \Sigma_{(a)}(\mathbf{s}_i))$.*

Proof. The proof proceeds by showing that if the tuning parameter λ is chosen correctly, then the penalty term vanishes for the relevant predictor groups and becomes infinite for the irrelevant predictor groups. \square

Since $\|\tilde{\beta}(\mathbf{s}_i)\|^\gamma = O_p\{(nh^2)^{-\gamma/2}\}$ and $h = O(n^{-1/6})$, in order for $\sqrt{na_n} \rightarrow 0$ and $\sqrt{nb_n} \rightarrow \infty$, we require that $\lambda = O(n^\alpha)$ where $\alpha \in (-\{1 + \gamma - \frac{\gamma}{6}\}/2, -1/2)$.

2. References

References

- Fan, J. and I. Gijbels (1996). *Local Polynomial Modeling and its Applications*. Chapman and Hall, London.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Boston MA.
- Hastie, T. and C. Loader (1993). Local regression: automatic kernel carpentry. *Statistical Science* 8, 120–143.

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.