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 $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n$, 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,

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

Thus, conditional on the design matrix $\boldsymbol{X}$, 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 (Härdle, 1990), 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 (Hastie and Loader, 1993).

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

$$\boldsymbol{Z}(\boldsymbol{s}_i) = (\boldsymbol{X} \quad L(\boldsymbol{s}_i)\, \boldsymbol{X} \quad M(\boldsymbol{s}_i)\, \boldsymbol{X}) \tag{3}$$

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

0.2. Estimation

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

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

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

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

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

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

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

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

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

This local likelihood can be maximized by weighted least squares

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

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

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

1. Asymptotics

1.1. Consistency

Theorem 1.1. *If $h\sqrt{n}a_n \xrightarrow{p} 0$ 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 $h = O(n^{-1/6})$.

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

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

We show the result:

$$
\begin{aligned}
Q\left\{\boldsymbol{\beta}(\boldsymbol{s}) + hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{u}\right\} - Q\left\{\boldsymbol{\beta}(\boldsymbol{s})\right\}&\\
= (1/2)&\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\left\{\boldsymbol{\beta}(\boldsymbol{s}) + hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{u}\right\}\right]^T \boldsymbol{W}(\boldsymbol{s})\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\left\{\boldsymbol{\beta}(\boldsymbol{s}) + hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{u}\right\}\right]\\
&+ n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}(\boldsymbol{s}) + hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{u}\|\\
&- (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\} - n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}(\boldsymbol{s})\|\\
= (1/2)&\boldsymbol{u}^T\left\{\frac{h^2}{n}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\boldsymbol{H}^{-1}\right\}\boldsymbol{u} - \boldsymbol{u}^T\left[hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right]\\
&+ n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_j(\boldsymbol{s}) + hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{u}\| - n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_j(\boldsymbol{s})\|\\
= (1/2)&\boldsymbol{u}^T\left\{\frac{h^2}{n}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\boldsymbol{H}^{-1}\right\}\boldsymbol{u} - \boldsymbol{u}^T\left[hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right]\\
&+ n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_j(\boldsymbol{s}) + hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{u}\| - n\sum_{j=1}^{p_0}\lambda_j\|\boldsymbol{\beta}_j(\boldsymbol{s})\|\\
\geqslant (1/2)&\boldsymbol{u}^T\left\{\frac{h^2}{n}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\boldsymbol{H}^{-1}\right\}\boldsymbol{u} - \boldsymbol{u}^T\left[hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right]\\
&+ n\sum_{j=1}^{p_0}\lambda_j(\|\boldsymbol{\beta}_j(\boldsymbol{s}) + hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{u}\| - \|\boldsymbol{\beta}_j(\boldsymbol{s})\|)\\
\geqslant (1/2)&\boldsymbol{u}^T\left\{\frac{h^2}{n}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})\boldsymbol{H}^{-1}\right\}\boldsymbol{u} - \boldsymbol{u}^T\left[hn^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{Z}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\}\right]\\
&- p_0 h\sqrt{n}a_n\|\boldsymbol{H}^{-1}\boldsymbol{u}\| \qquad (10)
\end{aligned}
$$

□

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

First term.. By Lemma 2 of Sun et al. (2014), $\frac{1}{n}\boldsymbol{Z}^T(\boldsymbol{s}_i)\boldsymbol{W}(\boldsymbol{s}_i)\boldsymbol{Z}(\boldsymbol{s}_i) \xrightarrow{p} \Omega$, so the first term in (10) converges to $\boldsymbol{u}^T\Omega\boldsymbol{u}$, a quadratic form in $\boldsymbol{u}$.

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

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

and so the second term of (10) is

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

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

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

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

So the expectation of (12) is:

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

And the variance of (12) is:

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

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

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

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

So the expectation of (15) is:

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

And the variance of (15) is:

$$
\begin{aligned}
&\mathrm{E}[\boldsymbol{H}^{-1}\,\{\boldsymbol{Z}^T(\boldsymbol{s})\}_i\,\{\boldsymbol{W}(\boldsymbol{s})\}_{ii}\,\varepsilon^2(\boldsymbol{s}_i)\,\{\boldsymbol{W}(\boldsymbol{s})\}_{ii}\,\{\boldsymbol{Z}(\boldsymbol{s})\}_i^T\,\boldsymbol{H}^{-1}] \\
&= \mathrm{E}\begin{pmatrix} \boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i) & h^{-1}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_1 & h^{-1}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_2 \\ h^{-1}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_1 & h^{-2}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_1^2 & h^{-2}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_1(\boldsymbol{s}_i-\boldsymbol{s} \\ h^{-1}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_2 & h^{-2}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_1(\boldsymbol{s}_i-\boldsymbol{s})_2 & h^{-2}\boldsymbol{X}(\boldsymbol{s}_i)\boldsymbol{X}^T(\boldsymbol{s}_i)(\boldsymbol{s}_i-\boldsymbol{s})_2^2 \end{pmatrix} \\
&\quad \times E\,\varepsilon^2(\boldsymbol{s}_i) \int K_h^2(\|\boldsymbol{s}-\boldsymbol{t}\|) f(\boldsymbol{t}) \partial \boldsymbol{t} \\
&= \begin{pmatrix} \boldsymbol{\Psi} & h^{-1}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_1 & h^{-1}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_2 \\ h^{-1}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_1 & h^{-2}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_1^2 & h^{-2}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_1(\boldsymbol{s}_i-\boldsymbol{s})_2 \\ h^{-1}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_2 & h^{-2}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_1(\boldsymbol{s}_i-\boldsymbol{s})_2 & h^{-2}\boldsymbol{\Psi}(\boldsymbol{s}_i-\boldsymbol{s})_2^2 \end{pmatrix} \times \sigma^2 \times h^{-2} f(\boldsymbol{s}) \nu_0 \\
&= 0 \qquad (17)
\end{aligned}
$$

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 $Q\left\{\boldsymbol{\beta}(\boldsymbol{s}_i)+n^{-1/2}\boldsymbol{u}\right\} >$ $Q\left\{\boldsymbol{\beta}(\boldsymbol{s}_i)\right\}$ is positive, which proves the result.

1.2. Selection

Theorem 1.2. *If* $\sqrt{n}a_n \xrightarrow{p} 0$ *and* $\sqrt{n}b_n \xrightarrow{p} \infty$ *then* $P\left\{\hat{\boldsymbol{\beta}}_{(b)}(\boldsymbol{s}_i)=0\right\} \to 1$.

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

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

$$Q\left\{\boldsymbol{\beta}(\boldsymbol{s}_i)\right\} = (1/2)\left\{\boldsymbol{Y}-\boldsymbol{Z}(\boldsymbol{s}_i)\boldsymbol{\beta}(\boldsymbol{s}_i)\right\}^T \boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{Y}-\boldsymbol{Z}(\boldsymbol{s}_i)\boldsymbol{\beta}(\boldsymbol{s}_i)\right\} + n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_p(\boldsymbol{s}_i)\| \tag{18}$$

Let $\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)\neq 0$. Then (18) is differentiable w.r.t. $\boldsymbol{\beta}_p(\boldsymbol{s}_i)$ and Q is maximized at

$$\begin{aligned}
0 &= \boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{Y}-\boldsymbol{X}_{(-p)}\hat{\boldsymbol{\beta}}_{-p}(\boldsymbol{s}_i)-\boldsymbol{X}_{(p)}\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)\right\} + n\lambda_p\frac{\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)}{\|\boldsymbol{\beta}_p(\boldsymbol{s}_i)\|} \\
&= \boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}_i)+\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}_i)-\boldsymbol{X}_{(-p)}\hat{\boldsymbol{\beta}}_{-p}(\boldsymbol{s}_i)-\boldsymbol{X}_{(p)}\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)\right\} + n\lambda_p\frac{\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)}{\|\boldsymbol{\beta}_p(\boldsymbol{s}_i)\|} \\
&= \boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}_i)\right\} + \boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{X}_{(-p)}\boldsymbol{\beta}_{-p}(\boldsymbol{s}_i)-\boldsymbol{X}_{(-p)}\hat{\boldsymbol{\beta}}_{-p}(\boldsymbol{s}_i)\right\} \\
&\qquad + \boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{X}_{(-p)}\boldsymbol{\beta}_{-p}(\boldsymbol{s}_i)-\boldsymbol{X}_{(p)}\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)\right\} + n\lambda_p\frac{\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)}{\|\boldsymbol{\beta}_p(\boldsymbol{s}_i)\|} \\
&= \sqrt{f(\boldsymbol{s}_i)h^2n^{-1}}\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}_i)\right\} + \sqrt{f(\boldsymbol{s}_i)h^2n^{-1}}\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\boldsymbol{X}_{(-p)}\left\{\boldsymbol{\beta}_{-p}(\boldsymbol{s}_i)-\hat{\boldsymbol{\beta}}_{-p}(\boldsymbol{s}_i)\right\} \\
&\qquad + \sqrt{f(\boldsymbol{s}_i)h^2n^{-1}}\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\boldsymbol{X}_{(-p)}\left\{\boldsymbol{\beta}_{-p}(\boldsymbol{s}_i)-\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)\right\} + \sqrt{f(\boldsymbol{s}_i)h^2n}\lambda_p\frac{\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)}{\|\boldsymbol{\beta}_p(\boldsymbol{s}_i)\|} \\
&= \sqrt{f(\boldsymbol{s}_i)h^2n^{-1}}\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}_i)\right\} + n^{-1}\left\{\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\boldsymbol{X}_{(-p)}\right\}\sqrt{f(\boldsymbol{s}_i)h^2n}\left\{\boldsymbol{\beta}_{-p}(\boldsymbol{s}_i)-\hat{\boldsymbol{\beta}}_{-p}(\boldsymbol{s}_i)\right\} \\
&\qquad + n^{-1}\left\{\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\boldsymbol{X}_{(-p)}\right\}\sqrt{f(\boldsymbol{s}_i)h^2n}\left\{\boldsymbol{\beta}_{-p}(\boldsymbol{s}_i)-\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)\right\} + \sqrt{f(\boldsymbol{s}_i)h^2n}\lambda_p\frac{\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}_i)}{\|\boldsymbol{\beta}_p(\boldsymbol{s}_i)\|}
\end{aligned} \tag{19}$$

From Lemma 2 of Sun et al. (2014), $n^{-1}\left\{\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\boldsymbol{X}_{(-p)}\right\} = O_p(1)$ and $n^{-1}\left\{\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\boldsymbol{X}_{(p)}\right\} = O_p(1)$. From Theorem 3 of Sun et al. (2014), we have that $\sqrt{f(\boldsymbol{s}_i)h^2n}\left\{\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}_i)-\hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s}_i)\right\} = O_p(1)$ and $\sqrt{f(\boldsymbol{s}_i)h^2n}\left\{\boldsymbol{\beta}_{(p)}(\boldsymbol{s}_i)-\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i)\right\} = O_p(1)$. So the second and third terms of the sum in (19) are $O_p(1)$. We showed in the proof of 1.1 that $\sqrt{f(\boldsymbol{s}_i)h^2n^{-1}}\boldsymbol{X}_{(p)}^T\boldsymbol{W}(\boldsymbol{s}_i)\left\{\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}_i)\right\} = O_p(1)$.

Because the first three terms of the sum in 19 are $O_p(1)$, for $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i)$ to be a solution, we must have that $\sqrt{f(\boldsymbol{s}_i)h^2n}\lambda_p \frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i)}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i)\|} = O_p(1)$.

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 $|\hat{\beta}_{(p),k}(\boldsymbol{s}_i)| = \max\{|\hat{\beta}_{(p),k'}(\boldsymbol{s}_i)| : 1 \leqslant k' \leqslant d_p\}$. And for this k, we have that $|\hat{\beta}_{(p),k}(\boldsymbol{s}_i)|/\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i)\| \geqslant 1/\sqrt{d_p} > 0$.

Now since $\sqrt{n}b_n \to \infty$, we have that $\sqrt{f(\boldsymbol{s}_i)h^2n}\lambda_p \frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i)}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i)\|}$ is unbounded and therefore dominates the $O_p(1)$ terms of the sum in (19). So for large enough n, $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i) \neq 0$ cannot maximize Q. □

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 $\sqrt{n}a_n \to 0$ and $\sqrt{n}b_n \to \infty$, then* $\sqrt{nh^2 f(\boldsymbol{s})}\left(\hat{\boldsymbol{\beta}}_{(a)}(\boldsymbol{s}_i) - \boldsymbol{\beta}_{(a)}(\boldsymbol{s}_i) - \frac{\kappa_2 h^2}{2\kappa_0}\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}_i) + \boldsymbol{\beta}_{vv}(\boldsymbol{s}_i)\}\right) \xrightarrow{d} N(0, \Sigma_{(a)}(\boldsymbol{s}_i))$.

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.

□

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

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.

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

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.