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 $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, which are distributed in a spatial domain $D \subset \mathbb{R}^2$ according to a density $f(\boldsymbol{s})$. For $i = 1, \ldots, n$, 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 $\varepsilon(\boldsymbol{s}_i)$, $i = 1, \ldots, n$ are independent. That is,

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

In order to simplify the notation, let $\boldsymbol{x}(\boldsymbol{s}_i) \equiv \boldsymbol{x}_i \equiv (1, x_{i1}, \ldots, x_{ip})'$, $\boldsymbol{\beta}(\boldsymbol{s}_i) \equiv \boldsymbol{\beta}_i \equiv (\beta_{i0}, \beta_{i1}, \ldots, \beta_{ip})'$, and $y(\boldsymbol{s}_i) \equiv y_i$. Equations (1) and (2) can be rewritten

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta}_i + \varepsilon_i \text{ and } \varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right). \tag{3}$$

Further, let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ and $\boldsymbol{y} = (y_1, \ldots, y_n)'$. 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

$$\boldsymbol{Z}_i = (\boldsymbol{X} \;\; L_i\boldsymbol{X} \;\; M_i\boldsymbol{X}) \tag{4}$$

where $\boldsymbol{X}$ is the unaugmented matrix of covariates, $L_i = \text{diag}\{s_{i',x} - s_{i,x}\}$ and $M_i = \text{diag}\{s_{i',y} - s_{i,y}\}$ 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:

$$\ell(\boldsymbol{\beta}_i) = -(1/2) \sum_{i'=1}^{n} \left\{ \log \sigma_i^2 + \sigma_i^{-2} \left(y_{i'} - \boldsymbol{z}_{i'}' \boldsymbol{\beta}_i \right)^2 \right\}. \tag{5}$$

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 $\boldsymbol{\beta}_i$ are estimated at $\boldsymbol{s}_i$ by the weighted likelihood

$$\mathcal{L}_i\left(\boldsymbol{\beta}_i\right) = \prod_{i'=1}^{n} \left[\left(2\pi\sigma_i^2\right)^{-1/2} \exp\left\{ -1/2\sigma_i^{-2} \left(y_{i'} - \boldsymbol{z}_{i'}' \boldsymbol{\beta}_i \right)^2 \right\} \right]^{w_{ii'}}, \tag{6}$$

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

$$\begin{aligned} w_{ii'} &= K_h(\delta_{ii'}) = h^{-2} K\left(h^{-1}\delta_{ii'}\right) \\ K(x) &= \begin{cases} (3/4)(1-x^2) & \text{if } \delta_{ii'} < h, \\ 0 & \text{if } \delta_{ii'} \geqslant h. \end{cases} \end{aligned} \tag{7}$$

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

$$\ell_i\left(\boldsymbol{\beta}_i\right) = -(1/2) \sum_{i'=1}^{n} w_{ii'} \left\{ \log \sigma_i^2 + \sigma_i^{-2} \left(y_{i'} - \boldsymbol{z}_{i'}' \boldsymbol{\beta}_i \right)^2 \right\}. \tag{8}$$

This local likelihood can be maximized by weighted least squares

$$\hat{\boldsymbol{\beta}}(\boldsymbol{s}_i) = \left(\boldsymbol{Z}^T \boldsymbol{W}_i \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^T \boldsymbol{W}_i \boldsymbol{Y}. \tag{9}$$

From (8), the maximum local likelihood estimate $\hat{\sigma}_i^2$ is:

$$\hat{\sigma}_i^2 = \left(\sum_{i'=1}^{n} w_{ii'} \right)^{-1} \sum_{i'=1}^{n} w_{ii'} \left(y_{i'} - \boldsymbol{z}_{i'}' \hat{\boldsymbol{\beta}}_i \right)^2 \tag{10}$$

1. Asymptotics

consistency:

Theorem 1.1. *If* $\sqrt{n} a_n \xrightarrow{p} 0$ *then* $\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i - \frac{\kappa_2 h_1^2}{2\kappa_0} \{ \boldsymbol{\beta}_{uu,i} + \boldsymbol{\beta}_{vv,i} = O_p(n^{-1/2} h^{-1} f(\boldsymbol{s}_i)^{-1})$

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

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

$$\liminf_n P\left\{\inf_{u\in\mathcal{R}:\|u\|=C} Q\left(\boldsymbol{\beta}_i + n^{-1/3}u\right) > Q\left(\boldsymbol{\beta}_i\right)\right\} > 1-\epsilon$$

We show the result:

$$\begin{aligned}
Q\left(\boldsymbol{\beta}_i + n^{-1/2}\boldsymbol{u}\right) - Q\left(\boldsymbol{\beta}_i\right) = (1/2)\left(\boldsymbol{Y}-\boldsymbol{Z}\left\{\boldsymbol{\beta}_i + n^{-1/2}\boldsymbol{u}\right\}\right)^T \boldsymbol{W}_i\left(\boldsymbol{Y}-\boldsymbol{Z}\left\{\boldsymbol{\beta}_i + n^{-1/2}\boldsymbol{u}\right\}\right) + n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_i + n^{-1/2}\boldsymbol{u}\| \\
- (1/2)\left(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\beta}_i\right)^T \boldsymbol{W}\left(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\beta}_i\right) + n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_i\| \\
= (1/2)\boldsymbol{u}^T\left(\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{W}_i\boldsymbol{Z}\right)\boldsymbol{u} - \boldsymbol{u}^T\left(\frac{1}{n^{-1/2}}\boldsymbol{Z}^T\boldsymbol{W}_i(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\beta}_i)\right) \\
+ n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_{ij} + n^{-1/2}\boldsymbol{u}\| - n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_{ij}\| \\
= (1/2)\boldsymbol{u}^T\left(\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{W}_i\boldsymbol{Z}\right)\boldsymbol{u} - \boldsymbol{u}^T\left(\frac{1}{n^{-1/2}}\boldsymbol{Z}^T\boldsymbol{W}_i(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\beta}_i)\right) \\
+ n\sum_{j=1}^{p}\lambda_j\|\boldsymbol{\beta}_{ij} + n^{-1/2}\boldsymbol{u}\| - n\sum_{j=1}^{p_0}\lambda_j\|\boldsymbol{\beta}_{ij}\| \\
\geqslant (1/2)\boldsymbol{u}^T\left(\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{W}_i\boldsymbol{Z}\right)\boldsymbol{u} - \boldsymbol{u}^T\left(\frac{1}{n^{-1/2}}\boldsymbol{Z}^T\boldsymbol{W}_i(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\beta}_i)\right) \\
+ n\sum_{j=1}^{p_0}\lambda_j(\|\boldsymbol{\beta}_{ij} + n^{-1/2}\boldsymbol{u}\| - \|\boldsymbol{\beta}_{ij}\|) \\
\geqslant (1/2)\boldsymbol{u}^T\left(\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{W}_i\boldsymbol{Z}\right)\boldsymbol{u} - \boldsymbol{u}^T\left(\frac{1}{n^{-1/2}}\boldsymbol{Z}^T\boldsymbol{W}_i(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\beta}_i)\right) + p_0(\sqrt{n}a_n) \qquad (11)
\end{aligned}$$

□

We'll consider the terms of the sum in (11) separately.

First term.. By Lemma 2 of **?**, $\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{W}_i\boldsymbol{Z} \xrightarrow{p} \Omega$, so the first term in 11 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 $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{i'} + \nabla\boldsymbol{\beta}_{i'}(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_{i'})$ for $i' = 1, \ldots, n$. So

$$\begin{aligned}
\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_i &= \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \boldsymbol{Z}_1^T(\boldsymbol{\beta}_1 + \nabla\boldsymbol{\beta}_1(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_1)) \\ \vdots \\ \boldsymbol{Z}_n^T(\boldsymbol{\beta}_n + \nabla\boldsymbol{\beta}_n(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_n)) \end{pmatrix} \\
&= \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \boldsymbol{m}_1 + \boldsymbol{Z}_1^T(\nabla\boldsymbol{\beta}_1(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_1)) \\ \vdots \\ \boldsymbol{m}_n + \boldsymbol{Z}_n^T(\nabla\boldsymbol{\beta}_n(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_n)) \end{pmatrix} \\
&= \boldsymbol{\varepsilon} - \begin{pmatrix} \boldsymbol{Z}_1^T(\nabla\boldsymbol{\beta}_1(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_1)) \\ \vdots \\ \boldsymbol{Z}_n^T(\nabla\boldsymbol{\beta}_n(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_n)) \end{pmatrix}
\end{aligned}$$

and so the second term of 11 is

$$\boldsymbol{u}^T \left(\frac{1}{n^{-1/2}} \boldsymbol{Z}^T \boldsymbol{W}_i (\boldsymbol{\varepsilon} - \begin{pmatrix} \boldsymbol{Z}_1^T(\nabla\boldsymbol{\beta}_1(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_1)) \\ \vdots \\ \boldsymbol{Z}_n^T(\nabla\boldsymbol{\beta}_n(\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_n)) \end{pmatrix}) \right)$$

which is $O_p(1)$.

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 $Q\left(\boldsymbol{\beta}_i + n^{-1/3}u\right) > Q\left(\boldsymbol{\beta}_i\right)$ is positive, which proves the result.

selection:
Theorem 1.2. *If* $\sqrt{n}a_n \xrightarrow{p} 0$ *then* $\hat{\boldsymbol{\beta}}(\boldsymbol{s}) - \boldsymbol{\beta}(\boldsymbol{s}) - \frac{\kappa_2 h_1^2}{2\kappa_0}\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{vv}(\boldsymbol{s})\} = O_p(n^{-1/2}h^{-1}f(\boldsymbol{s})^{-1})$

oracle property:

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.