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

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