Web-based Supplementary Material for "Local Adaptive Grouped Regularization and its Oracle Properties for Varying Coefficient Regression"

Wesley Brooks[a], Jun Zhu[b], Zudi Lu[c]

[a]*Department of Statistics, University of Wisconsin, Madison, WI 53706*

[b]*Department of Statistics and Department of Entomology, University of Wisconsin, Madison, WI 53706*

[c]*School of Mathematical Sciences, The University of Southampton, Highfield, Southampton UK*

Abstract

Varying coefficient regression is a flexible technique for modeling data where the coefficients are functions of some effect-modifying parameter, often time or location. While there are a number of methods for variable selection in a varying coefficient regression model, the existing methods mostly do global selection, which includes or excludes each covariate over the entire domain. Presented here is a new local adaptive grouped regularization (LAGR) method for local variable selection in spatially varying coefficient linear and generalized linear regression. LAGR selects the covariates that are associated with the response at any point in space, and simultaneously estimates the coefficients of those covariates by tailoring the adaptive group Lasso toward a local regression model with locally linear coefficient estimates. Oracle properties of the proposed method are established under local linear regression and local generalized linear regression. The finite sample properties of LAGR are assessed in a simulation study and for illustration, the Boston housing price data set is analyzed.

Keywords: adaptive Lasso, local generalized linear regression, local linear regression, regularization method, nonparametric, varying coefficient model

Email addresses: `wrbrooks@uwalumni.com` (Wesley Brooks), `jzhu@stat.wisc.edu` (Jun Zhu), `Z.Lu@soton.ac.uk` (Zudi Lu)

1. Lemmas

Lemma 1.

$$E\left[\sum_{i=1}^{n} q_1\left(\boldsymbol{Z}_i^T \boldsymbol{\zeta}(\boldsymbol{s}), Y_i\right) \boldsymbol{Z}_i K_h\left(\|\boldsymbol{s}-\boldsymbol{s}_i\|\right)\right] =$$

$$\begin{pmatrix} 2^{-1} n^{1/2} h^3 f(\boldsymbol{s}) \kappa_2 \boldsymbol{\Gamma}(\boldsymbol{s}) \left(\nabla_{uu}^2 \boldsymbol{\beta}(\boldsymbol{s}) + \nabla_{vv}^2 \boldsymbol{\beta}(\boldsymbol{s})\right)^T \\ \boldsymbol{0}_{2p} \end{pmatrix} + o_p\left(h^2 \boldsymbol{1}_{3p}\right)$$

and

$$Var\left[\sum_{i=1}^{n} q_1\left(\boldsymbol{Z}_i^T \boldsymbol{\zeta}(\boldsymbol{s}), Y_i\right) \boldsymbol{Z}_i K_h\left(\|\boldsymbol{s}-\boldsymbol{s}_i\|\right)\right] = f(\boldsymbol{s}) \mathrm{diag}\left\{\nu_0, \nu_2, \nu_2\right\} \otimes \boldsymbol{\Gamma}(\boldsymbol{s}) + o\left(1\right)$$

$$= \Lambda + o\left(1\right)$$

Proof. **Expectation**: For $j = 1, \ldots, p$, by a Taylor expansion of $\beta_j(\boldsymbol{s}_i)$ around $\boldsymbol{s}$,

$$\beta_j(\boldsymbol{s}_i) = \beta_j(\boldsymbol{s}) + \nabla\beta_j(\boldsymbol{s})(\boldsymbol{s}_i - \boldsymbol{s}) + (\boldsymbol{s}_i - \boldsymbol{s})^T \left\{\nabla^2\beta_j(\boldsymbol{s})\right\}(\boldsymbol{s}_i - \boldsymbol{s}) + o\left(h^2\right)$$

and thus, for $\boldsymbol{x} \in \mathbb{R}^p$,

$$\boldsymbol{x}_i^T \boldsymbol{\beta}(\boldsymbol{s}_i) = \sum_{j=1}^{p} x_{ij}\left[\beta_j(\boldsymbol{s}) + \nabla\beta_j(\boldsymbol{s})^T(\boldsymbol{s}_i - \boldsymbol{s}) + \tilde{\beta}_{ij}''\right] + o\left(h^2\right).$$

Letting $\boldsymbol{z}_i^T = \left\{(1, s_{i,1} - s_1, s_{i,2} - s_2) \otimes \boldsymbol{x}_i^T\right\}$ and $\boldsymbol{\zeta}(\boldsymbol{s}) = \left(\boldsymbol{\beta}(\boldsymbol{s})^T, \nabla_u\boldsymbol{\beta}(\boldsymbol{s})^T, \nabla_v\boldsymbol{\beta}(\boldsymbol{s})^T\right)^T$, we have that

$$\boldsymbol{x}_i^T \boldsymbol{\beta}(\boldsymbol{s}_i) - \boldsymbol{z}_i^T \boldsymbol{\zeta}(\boldsymbol{s}) = \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}_i'' + o\left(h^2\right)$$

$$= O_p\left(h^2\right).$$

By a Taylor expansion around $\boldsymbol{x}^T\boldsymbol{\beta}(\boldsymbol{s}_i)$, then,

$$q_1\left(\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}),\mu(\boldsymbol{s}_i,\boldsymbol{z}_i)\right) =q_1\left(\boldsymbol{x}_i^T\boldsymbol{\beta}(\boldsymbol{s}_i),\mu(\boldsymbol{s}_i,\boldsymbol{z})\right)$$
$$-q_2\left(\boldsymbol{x}_i^T\boldsymbol{\beta}(\boldsymbol{s}_i),\mu(\boldsymbol{s}_i,\boldsymbol{z})\right)\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}_i''$$
$$+o\left(h^2\right).$$

And by the definitions of $q_1(\cdot)$ and $q_2(\cdot)$, we have that

$$q_1\left(\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}),\mu(\boldsymbol{s}_i,\boldsymbol{z}_i)\right)=\rho(\boldsymbol{s}_i,\boldsymbol{z}_i)\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}_i''+o\left(h^2\right).$$

Now the expectation of Ω_n is

$$nE\left(\omega_i|\boldsymbol{Z}_i=\boldsymbol{z}_i,\boldsymbol{s}_i\right)=(1/2)\,\alpha_n\boldsymbol{z}_iq_1\left(\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}),\mu(\boldsymbol{s}_i,\boldsymbol{z}_i)\right)K\left(h^{-1}\|\boldsymbol{s}-\boldsymbol{s}_i\|\right)$$
$$=(1/2)\,\alpha_nh^2\boldsymbol{z}_i\left\{h^{-2}\rho(\boldsymbol{s}_i,\boldsymbol{z}_i)\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}_i''+o\left(\mathbf{1}_{3p}\right)\right\}K\left(h^{-1}\|\boldsymbol{s}-\boldsymbol{s}_i\|\right).$$

To facilitate a change of variables, we observe that $h^{-2}\tilde{\beta}_j''=\left(\frac{\boldsymbol{s}_i-\boldsymbol{s}}{h}\right)^T\left\{\nabla^2\beta_j(\boldsymbol{s})\right\}\left(\frac{\boldsymbol{s}_i-\boldsymbol{s}}{h}\right)$. Thus,

$$E\left(\omega_i|\boldsymbol{s}_i\right)=(1/2)\,\alpha_nh^2\left[\left(\begin{array}{c}1\\h^{-1}(s_{i,1}-s_1)\\h^{-1}(s_{i,2}-s_2)\end{array}\right)\otimes\left\{\boldsymbol{\Gamma}(\boldsymbol{s}_i)h^{-2}\tilde{\boldsymbol{\beta}}_i''\right\}+o\left(\mathbf{1}_{3p}\right)\right]K\left(h^{-1}\|\boldsymbol{s}-\boldsymbol{s}_i\|\right).$$

And, using the symmetry of the kernel function,

$$E\left(\omega_i\right)=(1/2)\alpha_nh^4f(\boldsymbol{s})\left(\begin{array}{c}\kappa_2\\h\kappa_3\\h\kappa_3\end{array}\right)\otimes\left[\boldsymbol{\Gamma}(\boldsymbol{s})\left\{\nabla^2_{uu}\boldsymbol{\beta}(\boldsymbol{s})+\nabla^2_{vv}\boldsymbol{\beta}(\boldsymbol{s})\right\}\right]+o\left(h^2\mathbf{1}_{3p}\right)$$

where $\{\nabla^2_{uu}\boldsymbol{\beta}(\boldsymbol{s}) + \nabla^2_{vv}\boldsymbol{\beta}(\boldsymbol{s})\} = (\nabla^2_{uu}\beta_1(\boldsymbol{s}) + \nabla^2_{vv}\beta_1(\boldsymbol{s}), \ldots, \nabla^2_{uu}\beta_p(\boldsymbol{s}) + \nabla^2_{vv}\beta_p(\boldsymbol{s}))^T$. Thus,

$$E\left(\Omega_n\right) = \begin{pmatrix} \alpha_n^{-1}2^{-1}h^2\kappa_2 f(\boldsymbol{s})\boldsymbol{\Gamma}(\boldsymbol{s})\left(\nabla^2_{uu}\boldsymbol{\beta}(\boldsymbol{s}) + \nabla^2_{vv}\boldsymbol{\beta}(\boldsymbol{s})\right)^T \\ \mathbf{0}_{2p} \end{pmatrix} + o_p\left(h^2\mathbf{1}_{3p}\right)$$

Variance: By the previous result, $E\left(\Omega_n\right) = O\left(h^2\right)$. Thus, $var\left(\Omega_n\right) \to E\left(\Omega_n^2\right)$, and since the observations are independent, $E\left(\Omega_n^2\right) = \sum_{i=1}^n E\left(\omega_i^2\right)$. And, by Taylor expansion around $\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}_i)$,

$$\begin{aligned} q_1^2\left(\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}), Y_i\right) =& q_1^2\left(\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}_i), Y_i\right) \\ &- q_1\left(\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}_i), Y_i\right) q_2\left(\boldsymbol{z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}_i), Y_i\right) \boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}_i'' \\ &+ o\left(h^2\right). \end{aligned}$$

Since $q_1\left(\cdot, \cdot\right)$ is the quasi-score function, it follows that

$$E\left(\omega_i^2 | \boldsymbol{Z}_i = \boldsymbol{z}_i, \boldsymbol{s}_i\right) = \alpha_n^2\rho(\boldsymbol{s}_i, \boldsymbol{z}_i)\boldsymbol{z}_i\boldsymbol{z}_i^T K\left(h^{-1}\|\boldsymbol{s} - \boldsymbol{s}_i\|\right) + o\left(h^2\right).$$

By the symmetry of the kernel function,

$$E\left(\omega_i^2\right) = n^{-1}f(\boldsymbol{s})\mathrm{diag}\left\{\nu_0, \nu_2, \nu_2\right\} \otimes \boldsymbol{\Gamma}(\boldsymbol{s}) + o\left(1\right).$$

Thus,

$$Var\left(\Omega_n\right) = f(\boldsymbol{s})\mathrm{diag}\left\{\nu_0, \nu_2, \nu_2\right\} \otimes \boldsymbol{\Gamma}(\boldsymbol{s}) + o\left(1\right).$$

$\square$

Lemma 2.

$$E\left[\sum_{i=1}^n q_2\left(\boldsymbol{Z}_i^T\boldsymbol{\zeta}(\boldsymbol{s}), Y_i\right)\boldsymbol{Z}_i\boldsymbol{Z}_i^T K_h\left(\|\boldsymbol{s} - \boldsymbol{s}_i\|\right)\right] = -f(\boldsymbol{s})\mathrm{diag}\left\{\kappa_0, \kappa_2, \kappa_2\right\} \otimes \boldsymbol{\Gamma}(\boldsymbol{s}) + o\left(1\right)$$

$$= -\Delta + o\left(1\right)$$

and

$$Var\left\{ \left(\sum_{i=1}^{n} q_2 \left(\mathbf{Z}_i^T \boldsymbol{\zeta}(\mathbf{s}), Y_i \right) \mathbf{Z}_i \mathbf{Z}_i^T K_h \left(\|\mathbf{s} - \mathbf{s}_i\| \right) \right)_{ij} \right\} = O\left(n^{-1} h^{-2} \right)$$

Proof. **Expectation**: The approach is similar to the proof of Lemma 1. By the Taylor expansion of $q_2\left(\mathbf{z}_i^T \boldsymbol{\zeta}(\mathbf{s}), \mu\left(\mathbf{s}_i, \mathbf{z}_i \right) \right)$ around $\mathbf{z}_i^T \boldsymbol{\zeta}(\mathbf{s}_i)$:

$$q_2\left(\mathbf{z}_i^T \boldsymbol{\zeta}(\mathbf{s}), \mu(\mathbf{s}_i, \mathbf{z}_i) \right) = q_2\left(\mathbf{z}_i^T \boldsymbol{\zeta}(\mathbf{s}_i), \mu(\mathbf{s}_i, \mathbf{z}_i) \right) + q_3\left(\mathbf{z}_i^T \boldsymbol{\zeta}(\mathbf{s}_i), \mu(\mathbf{s}_i, \mathbf{z}_i) \right) \left\{ \mathbf{z}_i^T \boldsymbol{\zeta}(\mathbf{s}) - \mathbf{z}_i^T \boldsymbol{\zeta}(\mathbf{s}_i) \right\}$$
$$= -\rho(\mathbf{s}_i, \mathbf{z}_i) + o\left(1 \right).$$

And by the same arguments as before

$$E\left(\delta_i | \mathbf{Z}_i = \mathbf{z}_i, \mathbf{s}_i \right) = -\alpha_n^2 \rho(\mathbf{s}_i, \mathbf{z}_i) \mathbf{z}_i \mathbf{z}_i^T K\left(h^{-1} \|\mathbf{s}_i - \mathbf{s}\| \right)$$

$$E\left(\delta_i | \mathbf{s}_i \right) = -\alpha_n^2 \begin{pmatrix} 1 \\ h^{-1}(s_{i,1} - s_1) \\ h^{-1}(s_{i,2} - s_2) \end{pmatrix} \begin{pmatrix} 1 \\ h^{-1}(s_{i,1} - s_1) \\ h^{-1}(s_{i,2} - s_2) \end{pmatrix}^T \otimes \boldsymbol{\Gamma}(\mathbf{s}_i) K\left(h^{-1} \|\mathbf{s}_i - \mathbf{s}\| \right)$$

$$E\left(\delta_i \right) = -n f\left(\mathbf{s} \right) \mathrm{diag}\left\{ \kappa_0, \kappa_2, \kappa_2 \right\} \otimes \boldsymbol{\Gamma}\left(\mathbf{s} \right) + o\left(n^{-1} \right)$$

Thus,

$$E\left(\Delta_n \right) = -f(\mathbf{s}) \mathrm{diag}\left\{ \kappa_0, \kappa_2, \kappa_2 \right\} \otimes \boldsymbol{\Gamma}(\mathbf{s}) + o\left(1 \right)$$

Variance: From the previous result, it follows that $\left\{ E\left(\delta_i \right) \right\}^2 = O\left(n^{-2} \right)$. By the definition of δ_i,

$$E\left(\delta_i^2 | \mathbf{Z}_i = \mathbf{z}_i, \mathbf{s}_i \right) =$$

$$\alpha_n^4 \mathbf{z}_i^T \mathbf{z}_i q_2^2(\mathbf{s}_i, \mathbf{z}_i) \begin{pmatrix} 1 \\ h^{-1}(s_{i,1} - s_1) \\ h^{-1}(s_{i,2} - s_2) \end{pmatrix} \begin{pmatrix} 1 \\ h^{-1}(s_{i,1} - s_1) \\ h^{-1}(s_{i,2} - s_2) \end{pmatrix}^T \mathbf{z}_i \mathbf{z}_i^T K^2\left(h^{-1} \|\mathbf{s}_i - \mathbf{s}\| \right) + o\left(1 \right)$$

And it follows that $E\left(\delta_i^2 \right) = O\left(n^{-1} \alpha_n^2 \right)$, and $Var\left(\Delta_n \right) = O\left(\alpha_n^2 \right)$. $\qquad \square$