# Meet this GAL: Geographically weighted regression with the Adaptive LASSO

Wesley Brooks

---

## 1. Introduction

Varying-coefficients regression (Hastie and Tibshirani, 1993) is a technique used in spatial statistics to model a non-stationary process. Geographically Weighted Regression (GWR) (Fotheringham et al., 2002) is a method of fitting varying-coefficients regression models for spatial data that uses kernel-weighted regression with weights based on the distance between observation locations. The presentation of GWR in Fotheringham et al. (2002) follows the development of local likelihood in Loader (1999).

GWR can be thought of as a kernel smoother for regression coefficients, and hence GWR coefficient estimates are likely to exhibit bias near the boundary of the region being modeled (Hastie and Loader, 1993). Modeling the coefficient surface as locally linear rather than locally constant (by including coefficient-by-geographic-location interactions) can reduce this boundary-effect bias (Hastie and Loader, 1993). Adding these interactions to the GWR model is analogous to a transition from kernel smoothing to local regression, and was introduced in Wang et al. (2008).

Some recent research has focused on variable selection in varying-coefficients models. In the context of varying-coefficients regression models, global variable selection (in which one compares the hypothesis that the coefficient on a given variable is zero everywhere against the hypothesis that

the coefficient is nonzero somewhere) is distinguished from local variable selection (in which one compares the hypothesis that the coefficient on a given variable is zero at a given location against the hypothesis that the coefficient at that location is nonzero). Global variable selection for models where the varying coefficients are estimated using splines is addressed in Fan and Zhang (1999) for response variables that belong to an exponential-family distribution (as in the generalized linear model), and in Wang et al. (2008) for models with repeated measurements. Antoniadas et al. (2012) estimates the coefficient functions with P-splines, and then uses the nonnegative garrote of Breiman (1995) to do local variable selection by selecting P-spline bases.

The geographically-weighted LASSO of Wheeler (2009) is used for local variable selection in GWR models.

## 2. Geographically-weighted regression models

### *2.1. Model*

Consider $n$ data observations, made at locations $s_1, \ldots, s_n$. For $i = 1, \ldots, n$, let $y(s_i)$ and $\boldsymbol{x}(s_i)$ be the univariate outcome of interest, and a $(p+1)$-variate vector of covariates measured at location $s_i$, respectively. At each location $s_i$, assume that the outcome is related to the covariates by a linear model with coefficients $\boldsymbol{\beta}_i(s_i)$ that may be spatially-varying.

$$y(s_i) = \boldsymbol{x}'(s_i)\boldsymbol{\beta}(s_i) + \epsilon(s_i) \tag{1}$$

Further assume that the error term $\epsilon(s)$ is normally distributed with zero mean and a possibly spatially-varying variance $\sigma^2(s)$

$$\epsilon(s_i) \sim \mathcal{N}\left(0, \sigma^2(s_i)\right) \tag{2}$$

In order to simplify the notation, let subscripts denote the values of data or parameters at the locations where data is observed. Thus, $\boldsymbol{x}(s_i) \equiv \boldsymbol{x}_i \equiv (1, x_{i1}, \ldots, x_{ip})'$, $\boldsymbol{\beta}(s_i) \equiv \boldsymbol{\beta}_i \equiv (\beta_{i0}, \beta_{i1}, \ldots, \beta_{ip})'$, $y(s_i) \equiv y_i$, and $\sigma^2(s_i) \equiv \sigma_i^2$. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ and $\boldsymbol{Y} = (y_1, \ldots, y_n)'$. Now equations 1 - 2 can be rewritten

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta}_i + \epsilon_i \tag{3}$$

$$\epsilon_i \sim \mathcal{N}\left(0, \sigma_i^2\right) \tag{4}$$

Assume that, given the covariates $\boldsymbol{X}$, observations of the output at different locations are statistically independent of each other. Then the total log-likelihood of the observed data is the sum of the log-likelihood of each individual observation.

$$\ell\left(\boldsymbol{\beta}\right) = -\frac{1}{2} \sum_{i=1}^{n} \left[ \log\left(2\pi\sigma_i^2\right) + \sigma_i^{-2} \left(y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_i\right)^2 \right] \tag{5}$$

With $n$ observations and $n \times (p+1)$ free parameters, the model is overdetermined so it is not possible to directly maximize the total likelihood. To effectively reduce the number of parameters, assume that the spatially-varying coefficients $\boldsymbol{\beta}(s)$ are *smoothly* varying, and use a kernel smoother to make pointwise estimates of the coefficients by maximizing the local likelihood. In the setting of spatial data and with the kernel smoother based on the physical distance between observation locations, this method is called geographically-weighted regression (GWR).

### *2.2. Geographically-weighted regression*

Geographically-weighted regression estimates the value of the coefficient surface $\boldsymbol{\beta}(s)$ at each location $s_i$. Assume for now that there are known weights $w_{ii'}$ based on the distance $\|s_i - s_{i'}\|$ between locations $s_i$ and $s_{i'}$ for all $i, i'$.

Coefficient estimation is done by maximizing the local likelihood at each location (Fotheringham et al., 2002).

$$L_i\left(\boldsymbol{\beta}_i\right) = \prod_{i'=1}^{n} \left\{ \left(2\pi\sigma_i^2\right)^{-\frac{1}{2}} \exp\left( -\frac{1}{2} \sigma_i^{-2} \left[ y_{i'} - \boldsymbol{x}_{i'}' \boldsymbol{\beta}_i \right]^2 \right) \right\}^{w_{ii'}} \tag{6}$$

$$\ell_i(\boldsymbol{\beta}_i) \quad \propto \quad -\frac{1}{2}\sum_{i'=1}^{n} w_{ii'}\left\{\log\sigma_i^2 + \sigma_i^{-2}\left(y_{i'} - \boldsymbol{x}_{i'}'\boldsymbol{\beta}_i\right)^2\right\} \tag{7}$$

The first and second derivatives of the local log-likelihood are

$$\left\{\frac{\partial \ell_i}{\partial \boldsymbol{\beta}_i}\right\}_j = \sum_{i'=1}^{n}\left\{x_{i'j} w_{ii'}\sigma_i^{-2}\left(y_{i'} - \boldsymbol{x}_{i'}'\boldsymbol{\beta}_i\right)\right\} \tag{8}$$

$$\left\{\frac{\partial^2 \ell_i}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i'}\right\}_{j,k} = -\sum_{i'=1}^{n}\left\{x_{i'j}x_{i'k} w_{ii'}\sigma_i^{-2}\right\} \tag{9}$$

So the observed Fisher information in the locally weighted sample is

$$\boldsymbol{\mathcal{J}}_i = \sigma_i^{-2}\begin{pmatrix} \sum_{i'=1}^{n} w_{ii'}x_{i'1}^2 & \dots & \sum_{i'=1}^{n} w_{ii'}x_{i'1}x_{i'p} \\ \vdots & \ddots & \vdots \\ \sum_{i'=1}^{n} w_{ii'}x_{i'p}x_{i'1} & \dots & \sum_{i'=1}^{n} w_{ii'}x_{i'p}^2 \end{pmatrix} \tag{10}$$

$$= \sigma_i^{-2}\sum_{i'=1}^{n} w_{ii'}\begin{pmatrix} x_{i'1}^2 & \dots & x_{i'1}x_{i'p} \\ \vdots & \ddots & \vdots \\ x_{i'p}x_{i'1} & \dots & x_{i'p}^2 \end{pmatrix} \tag{11}$$

$$= \sigma_i^{-2}\sum_{i'=1}^{n} w_{ii'}\boldsymbol{x}_{i'}\boldsymbol{x}_{i'}' \tag{12}$$

The form of the observed Fisher information suggests that the information in the data $\boldsymbol{x}_{i'}$ about the coefficients at location $s_i$ is proportional to the weight $w_{ii'}$.

At each location $s_i$, the ordinary geographically-weighted regression estimator minimizes the objective function:

$$\sum_{i'=1}^{n} w_{ii'}\left(y_{i'} - \boldsymbol{x}_{i'}'\boldsymbol{\beta}_i\right)^2 \tag{13}$$

Letting the weight matrix $\boldsymbol{W}_i$ be

$$\boldsymbol{W}_i = \begin{pmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{pmatrix} \tag{14}$$

estimation of the ordinary geographically-weighted regression coefficient surface is by weighted least squares:

$$\hat{\boldsymbol{\beta}}_{i,\mathrm{GWR}} = \left(\boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{Y} \tag{15}$$

*2.3. Smoothing kernel*

The bisquare kernel function is used to generate geographic weights based on the distance between observation locations. For estimating the value of the coefficient surface at location $s_i$, the weight given to the observation at location $s_{i'}$ is

$$w_{ii'} = \begin{cases} \left(1 - \left[\mathrm{bw}^{-1}\left\|s_i - s_{i'}\right\|\right]^2\right)^2 & \text{if } \left\|s_i - s_{i'}\right\| < \mathrm{bw} \\ 0 & \text{if } \left\|s_i - s_{i'}\right\| \geqslant \mathrm{bw} \end{cases} \tag{16}$$

where bw is the kernel bandwidth.

## 3. Model selection and shrinkage

Traditional GWR relies on *a priori* model selection to decide which variables should be included in the model. In the context of ordinary least squares regression, regularization methods such as the Adaptive LASSO (Zou, 2006) have been shown to have appealing properties for automating variable selection, sometimes including the "oracle" property of asymptotically selecting exactly the correct variables for inclusion in a regression model.

The Adaptive LASSO is applied to GWR by first multiplying the design matrix $\boldsymbol{X}$ by $\boldsymbol{W}_i$, the diagonal matrix of geographic weights centered at $s_i$. Since some of the weights $w_{ii'}$ may be zero, the matrix $\boldsymbol{W}_i\boldsymbol{X}$ is not of full rank. The matrices $\boldsymbol{Y}_i^*$, $\boldsymbol{X}_i^*$, and $\boldsymbol{W}_i^*$ are formed by dropping the rows of $\boldsymbol{X}$ and $\boldsymbol{W}_i$ that correspond to observations with zero weight in the regression model at location $s_i$. Now, letting $\boldsymbol{U}_i^* = \boldsymbol{W}_i^*\boldsymbol{X}_i^*$ and $\boldsymbol{V}_i^* = \boldsymbol{W}_i^*\boldsymbol{Y}_i^*$, we seek the coefficients $\boldsymbol{\beta}_i$ of the regression model:

$$\boldsymbol{V}_i^* = \boldsymbol{U}_i^*\boldsymbol{\beta}_i + \epsilon \tag{17}$$

To apply the Adaptive LASSO for estimating these regression coefficients, each column of $\boldsymbol{U}_i^*$ is centered around zero and rescaled to have an $L_2$-norm of one. Let $\tilde{\boldsymbol{U}}_i^*$ be the centered-and-scaled version of $\boldsymbol{U}_i^*$. Adaptive weights are calculated using the OLS regression coefficients $\boldsymbol{\gamma}_i^*$ via ordinary least squares (OLS):

$$\boldsymbol{\gamma}_i^* = \left(\tilde{\boldsymbol{U}}_i^{*\prime}\tilde{\boldsymbol{U}}_i^*\right)^{-1}\tilde{\boldsymbol{U}}_i^{*\prime}\boldsymbol{V}_i^* \tag{18}$$

Now a final scaling step is done: for $j = 1, \ldots, p$, the $j^{\text{th}}$ column of $\tilde{\boldsymbol{U}}_i^*$ is multiplied by $(\gamma_i^*)_j$, the corresponding coefficient from the regression equation in (18). Call this rescaled matrix $\breve{\boldsymbol{U}}_i^*$.

Finally, the Adaptive LASSO coefficient estimates at location $s_i$ are found by using the `lars` algorithm to model $\boldsymbol{V}_i^*$ as a function of $\breve{\boldsymbol{U}}_i^*$

### *3.1. Tuning parameter selection*

The final task is to select the LASSO tuning parameter. Wheeler (2009) proposed selecting the tuning parameter for the LASSO at location $s_i$ to minimize the jackknife prediction error $|y_i - \hat{y}_i^{(i)}|$,

which means that the coefficients can only be estimated at the locations where data has been observed. we propose to use the local AIC to select the tuning parameter, which allows coefficients to be estimated at any location where the local likelihood can be calculated. The local AIC is calculated by adding a penalty to the local likelihood, with the sum of the weights around $s_i$, $\sum_{i'=1}^{n} w_{ii'}$ playing the role of the sample size and the number of nonzero coefficients in $\boldsymbol{\beta}_i$ playing the role of the "degrees of freedom" ($\mathrm{df}_i$) (Zou et al., 2007).

The objective minimized by the geographically-weighted lasso (GWL) is:

$$\sum_{i'=1}^{n} w_{ii'} \left(y_{i'} - \boldsymbol{x}'_{i'} \boldsymbol{\beta}_i\right)^2 + \sum_{j=1}^{p} \lambda_{ij} \beta_{ij} \tag{19}$$

Where $\lambda_{ij}, j = 1, \ldots, p$ are penalties from the Adaptive LASSO (Zou, 2006). Taking the derivatives with respect to $\beta$ and setting to zero, we see that

$$\hat{\boldsymbol{\beta}}_{i,\mathrm{GWL}} = \left(\boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{Y} - \frac{1}{2}\left(\boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{X}\right)^{-1} \boldsymbol{\lambda}_i \tag{20}$$

$$\hat{y}_i = \boldsymbol{x}_i\hat{\boldsymbol{\beta}}_{i,\mathrm{GWL}} = \boldsymbol{x}_i \left(\boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{Y} - \frac{1}{2}\boldsymbol{x}_i \left(\boldsymbol{X}'\boldsymbol{W}_i\boldsymbol{X}\right)^{-1} \boldsymbol{\lambda}_i \tag{21}$$

So unlike in the case of ordinary geographically-weighted regression, the fitted values $\hat{\boldsymbol{Y}}$ are not a linear combination of the observations $\boldsymbol{Y}$. Because GWL is not a linear smoother the AIC and confidence intervals as calculated in Fotheringham et al. (2002) are not accurate for the GWL (Zou, 2006). The local AIC ($\mathrm{AIC}_{\mathrm{loc}}$) is minimized to select the adaptive lasso tuning parameter.

$$\mathrm{AIC}_{\mathrm{loc}} = \sum_{i'=1}^{n} w_{ii'} \hat{\sigma}_i^{-2} \left(y_{i'} - \boldsymbol{x}'_{i'}\hat{\boldsymbol{\beta}}_i\right)^2 + 2\mathrm{df}_i \tag{22}$$

Where the estimated local variance $\hat{\sigma}_i^2$ is the variance estimate from the unpenalized local model (Zou et al., 2007). The Maximum-Likelihood Estimate (MLE) of $\sigma_i^2$ is found by differentiating the local likelihood with respect to $\sigma_i^2$:

$$\left.\frac{\partial \ell_i}{\partial \sigma_i^2}\right|_{\hat{\beta}_i} = -\frac{1}{2}\sum_{i'=1}^{n} w_{ii'}\left\{\left(\sigma_i^2\right)^{-1} - \left(\sigma_i^2\right)^{-2}\left(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_i\right)^2\right\} \tag{23}$$

$$\hat{\sigma}_i^2 = \left(\sum_{i'=1}^{n} w_{ii'}\right)^{-1}\sum_{i'=1}^{n}\left(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_i\right) \tag{24}$$

*3.2. Bandwidth selection*

The bandwidth is selected to minimize the total AIC ($\mathrm{AIC}_{\mathrm{tot}}$). Because of the kernel weights and the application of the Adaptive LASSO, the sample size and degrees of freedom are different at each location. The total AIC is found by taking the sum over all of the observed data:

$$\mathrm{AIC}_{\mathrm{tot}} = \sum_{i=1}^{n}\left\{\hat{\sigma}_i^{-2}\left(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_i\right)^2 + \log\hat{\sigma}_i^2 + 2\mathrm{df}_i\left(\sum_{i'=1}^{n} w_{ii'}\right)^{-1}\right\} \tag{25}$$

This is different from the formulas for the AIC as proposed in **??** and **??**. The reason is that the basic GWR estimator is linear, so the degrees of freedom can be approximated using the trace of the "hat" matrix. The GAL, though, is not a linear estimator so some...

The bandwidth that minimizes (25) is found by a line search.

*3.3. Confidence interval estimation*

Confidence intervals for the GAL's coefficient estimates can be calculated either by the bootstrap (Efron and Tibshirani, 1986) or by using the variables selected by the Adaptive LASSO in a weighted least squares model. To compute coefficient confidence intervals via the bootstrap, the observations with non-zero geographic weights are resampled uniformly with replacement for each of $n_B$ bootstrap replicates. For each bootstrap replicate, the GWL is used to estimate regression coefficients. The local likelihood of the bootstrap replicates may be different from that of the original sample, so the adaptive lasso tuning parameter may differ for each bootstrap replicate. Since the GWL is

applied independently to each bootstrap replicate, the variables selected by GWL may be different for each replicate. The, e.g., 95% confidence interval for each regression coefficient is then the (2.5, 97.5) percentiles of the coefficient estimates from the bootstrap replicates.

### *3.3.1. Normal approximation-based confidence interval*

A third way to estimate the coefficient confidence intervals is to use the GWL for variable selection only and then to use GWR to calculate a confidence interval based on the assumption of an independent, identically distributed, Gaussian error structure. In this case, the standard error of the regression coefficients is

$$\hat{\mathrm{se}}_{\beta_i} = \left(\tilde{\boldsymbol{X}}_i' \boldsymbol{W}_i \tilde{\boldsymbol{X}}_i\right)^{-1} \tilde{\boldsymbol{X}}_i' \boldsymbol{W}_i \boldsymbol{Y} \tag{26}$$

where $\tilde{\boldsymbol{X}}_i$ is the model matrix including only those variables that are selected by GWL at location $i$.

### *3.3.2. Bootstrap confidence interval*

Unshrunk coefficient estimates are found by using the GWL at each location for variable selection only and then estimating the coefficients for the selected variables by GWR. An unshrunk bootstrap confidence interval is found by estimating the unshrunk coefficients for each of the $n_B$ bootstrap replicates and then calculating the percentiles as above.

## 4. Simulation

### *4.1. Simulation setup*

A simulation study was conducted to assess the finite-sample properties of the method described in Sections **??**-3. Data was simulated on $[0, 1] \times [0, 1]$, which was divided into a $30 \times 30$ grid. Each

of the $p = 5$ covariates was simulated by a Gaussian random field with mean zero and exponential covariance $Cov\left(Z_j(s_i), Z_j(s_{i'})\right) = \sigma^2 \exp\left(-\tau^{-1}\|s_i - s_{i'}\|\right)$ where $\sigma^2 = 1$ is the variance and $\tau$ is a range parameter. Correlation was induced between the covariates by multiplying the $\boldsymbol{Z}$ matrix by the Cholesky decomposition of the covariance matrix $\Sigma = \boldsymbol{R}'\boldsymbol{R}$. The covariance matrix is a $5 \times 5$ matrix that has ones on the diagonal and $\rho$ for all off-diagonal entries, where $\rho$ is the between-covariate correlation.

The simulated response is $y_i = \boldsymbol{x}_i\boldsymbol{\beta}_i + \epsilon_i$ for $i = 1, \ldots, 900$. The simulated data included the output $y$ and five covariates $x_1, \ldots, x_5$. The true data-generating model used only $x_1$, so $x_2, \ldots, x_5$ are included to test the variable-selection properties of GWL. The coefficient surface of $\beta_1$ is described by the "step" function:

$$\beta_1(s) = \begin{cases} 0 & \text{if } s_y < 0.4 \\ 5(s_y - 0.4) & \text{if } 0.4 \leqslant s_y < 0.6 \\ 1 & \text{o.w.} \end{cases} \tag{27}$$

.

In order to evaluate the performance of GWL under a range of conditions, the data was simulated under 18 different settings for each type of $\beta_1$ (Table **??**): high (0.1) and low (0.03) levels of the autoregression range parameter $\tau$ for the Gaussian random fields used to generate the covariates $\boldsymbol{X}_1(s), \ldots, \boldsymbol{X}_5(s)$; three levels (0, 0.5, 0.8) of between-covariate correlation $\rho$; and three levels (0, 0.03, 0.1) of the autoregression range parameter $\tau$ for the Gaussian random field used to generate the error term $\epsilon(s)$. Each case was simulated 100 times.

*4.2. Simulation results*

Results of the simulation experiment were summarized to asses the consistency in selection and estimation, as well as the coverage properties of the confidence intervals.

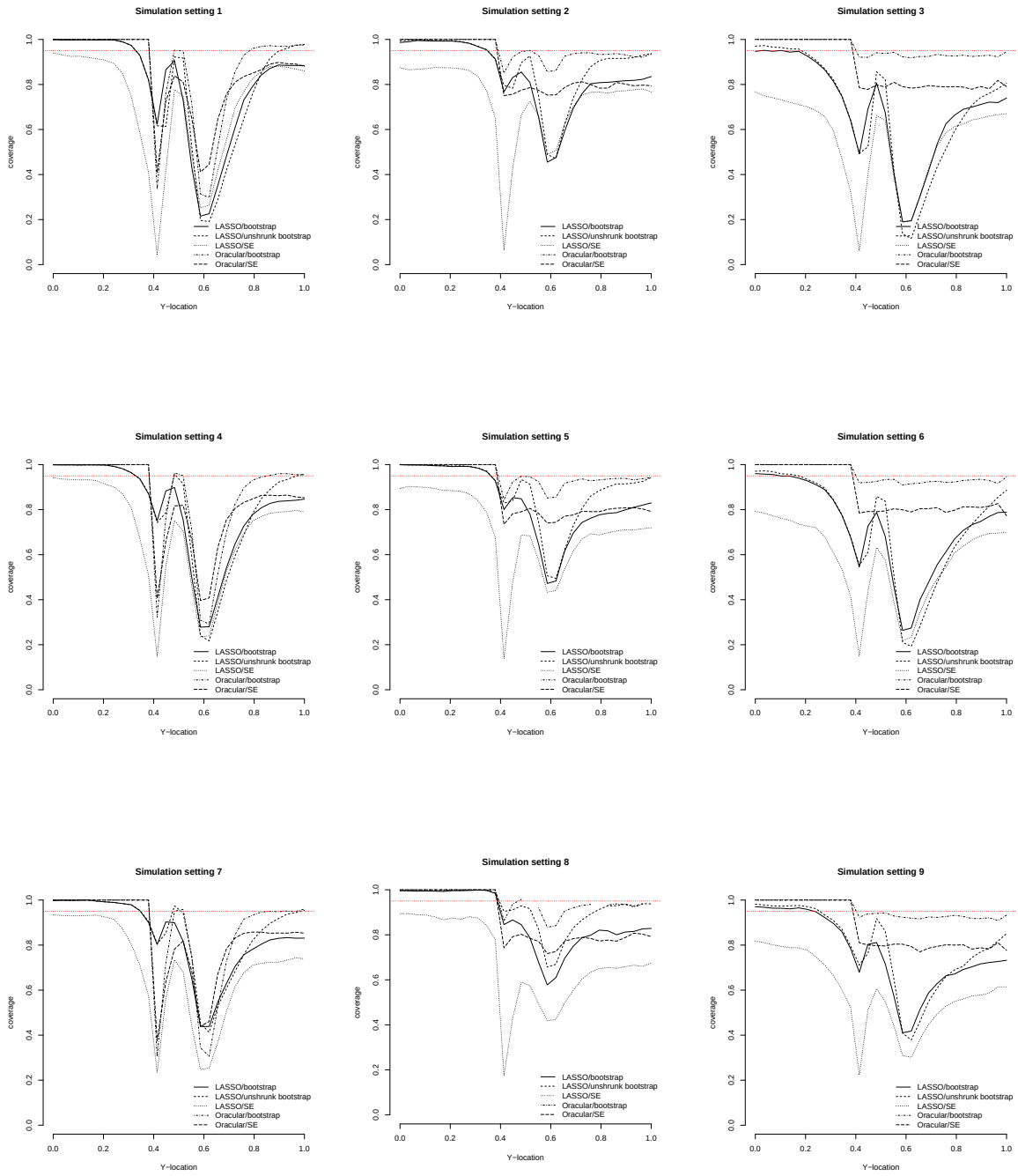
The data-generating process was simulated on the $30 \times 30$ grid spanning $[0, 1] \times [0, 1]$. Bootstrap confidence intervals were estimated with $n_B = 101$.

| | tau | rho | sigma.tau |
|---|---|---|---|
| 1 | 0.03 | 0.00 | 0.00 |
| 2 | 0.03 | 0.00 | 0.03 |
| 3 | 0.03 | 0.00 | 0.10 |
| 4 | 0.03 | 0.50 | 0.00 |
| 5 | 0.03 | 0.50 | 0.03 |
| 6 | 0.03 | 0.50 | 0.10 |
| 7 | 0.03 | 0.80 | 0.00 |
| 8 | 0.03 | 0.80 | 0.03 |
| 9 | 0.03 | 0.80 | 0.10 |
| 10 | 0.10 | 0.00 | 0.00 |
| 11 | 0.10 | 0.00 | 0.03 |
| 12 | 0.10 | 0.00 | 0.10 |
| 13 | 0.10 | 0.50 | 0.00 |
| 14 | 0.10 | 0.50 | 0.03 |
| 15 | 0.10 | 0.50 | 0.10 |
| 16 | 0.10 | 0.80 | 0.00 |
| 17 | 0.10 | 0.80 | 0.03 |
| 18 | 0.10 | 0.80 | 0.10 |

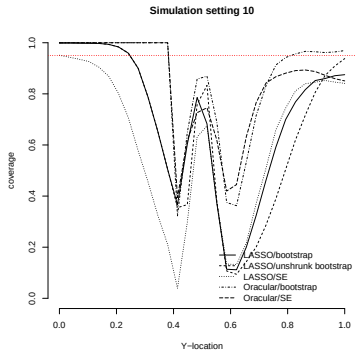
Table 1: Simulation parameters for each setting.

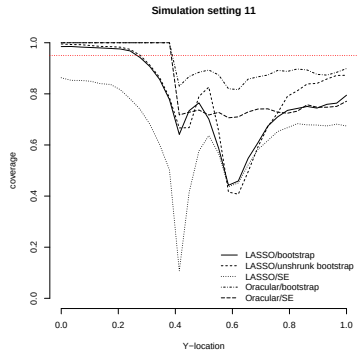
*4.3. Simulation results*

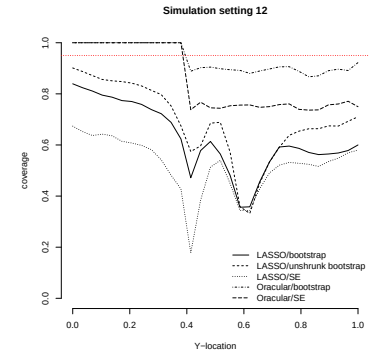
Results of the simulation experiment were summarized to asses the consistency in selection and estimation, as well as the coverage properties of the confidence intervals.

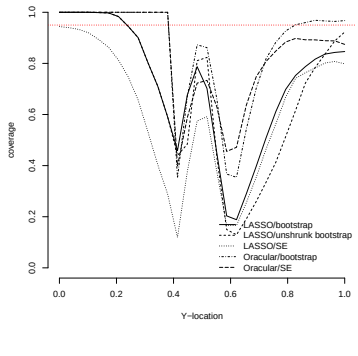

Simulation setting 1
Simulation setting 2
Simulation setting 3
Simulation setting 4
Simulation setting 5
Simulation setting 6
Simulation setting 7
Simulation setting 8
Simulation setting 9
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

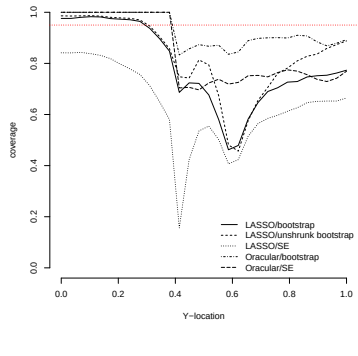
(a) A tiger

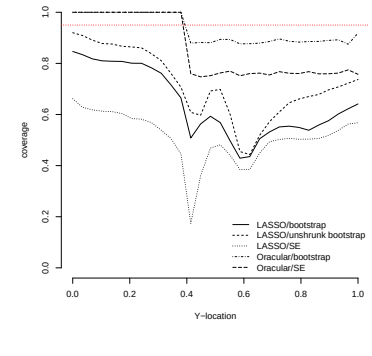

Simulation setting 10
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

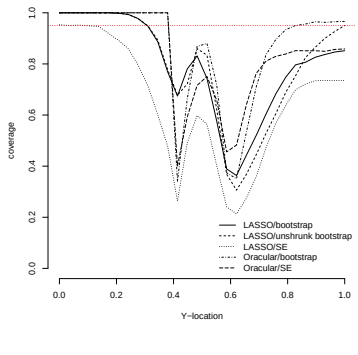

Simulation setting 11
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

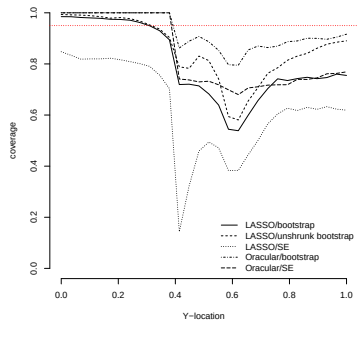

Simulation setting 12
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

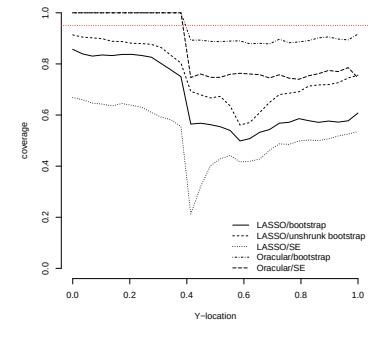

Simulation setting 13
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

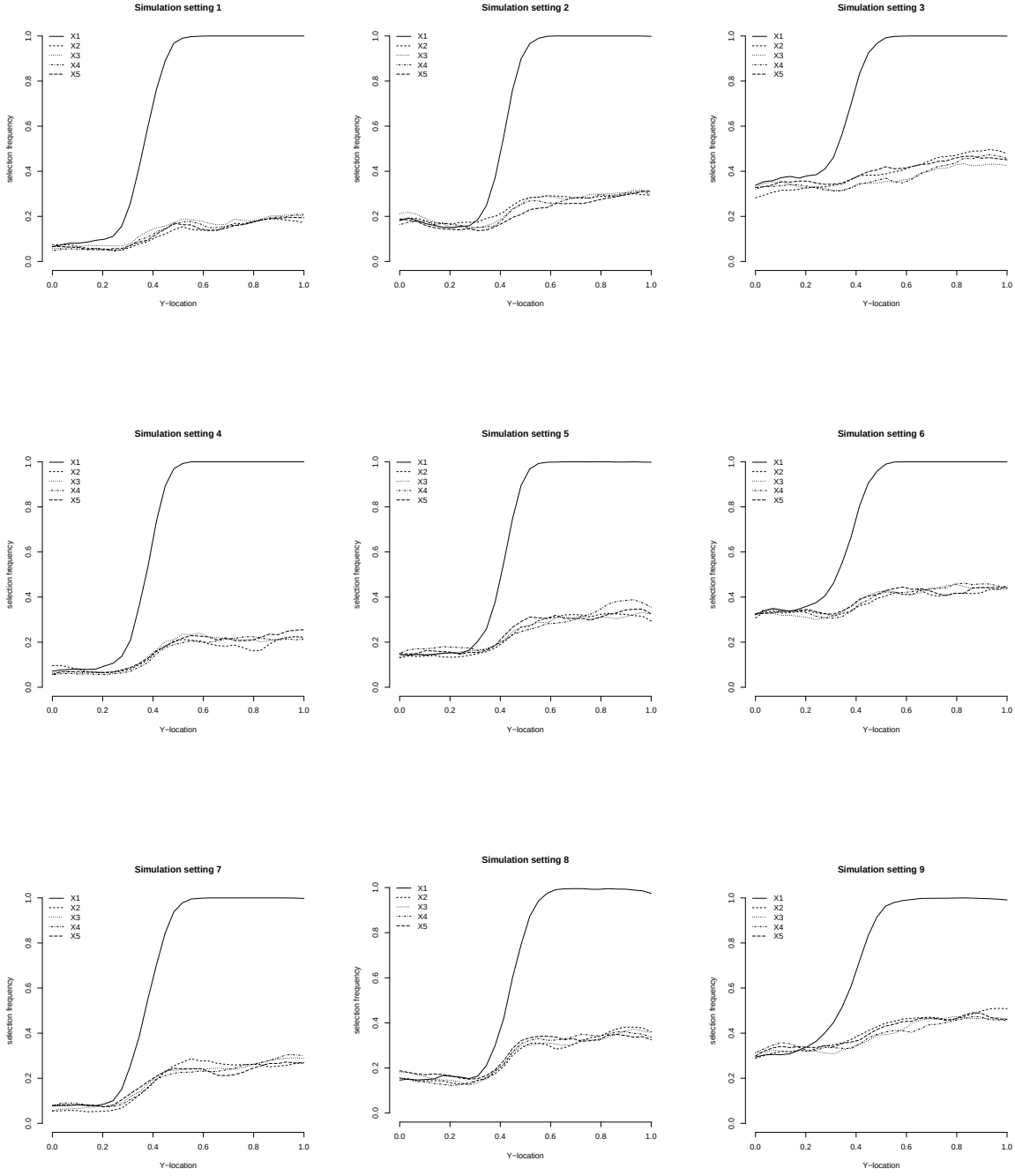

Simulation setting 14
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE


Simulation setting 15
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

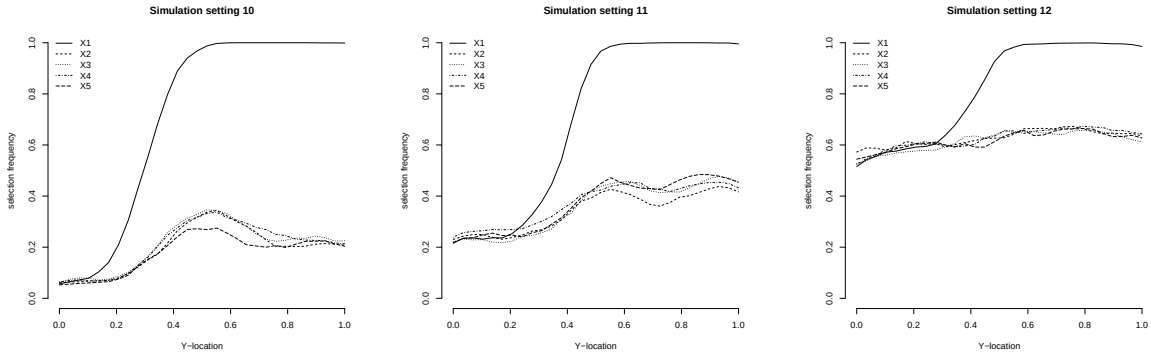

Simulation setting 16
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

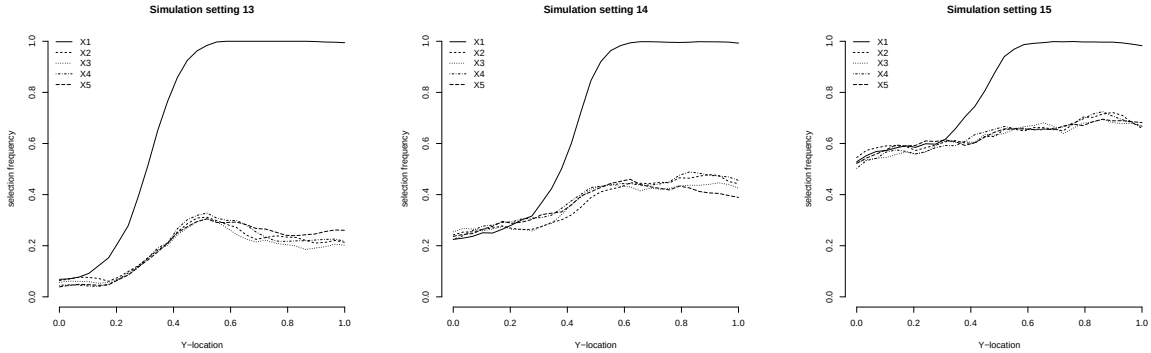

Simulation setting 17
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE

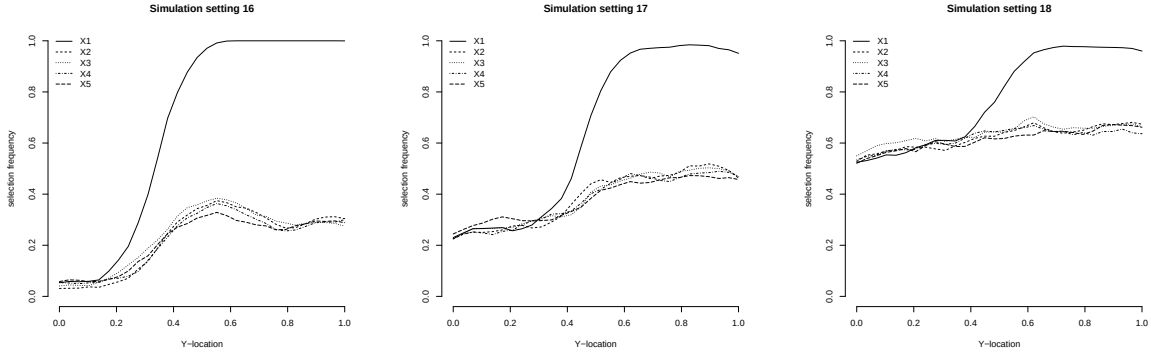

Simulation setting 18
coverage
Y−location
LASSO/bootstrap
LASSO/unshrunk bootstrap
LASSO/SE
Oracular/bootstrap
Oracular/SE


Simulation setting 1
Simulation setting 2
Simulation setting 3
Simulation setting 4
Simulation setting 5
Simulation setting 6
Simulation setting 7
Simulation setting 8
Simulation setting 9
X1
X2
X3
X4
X5
selection frequency
Y−location
0.0
0.2
0.4
0.6
0.8
1.0

(b) A tiger


Simulation setting 10
Simulation setting 11
Simulation setting 12
X1
X2
X3
X4
X5
selection frequency
Y−location
0.0
0.2
0.4
0.6
0.8
1.0


Simulation setting 13
Simulation setting 14
Simulation setting 15
X1
X2
X3
X4
X5
selection frequency
Y−location
0.0
0.2
0.4
0.6
0.8
1.0


Simulation setting 16
Simulation setting 17
Simulation setting 18
X1
X2
X3
X4
X5
selection frequency
Y−location
0.0
0.2
0.4
0.6
0.8
1.0

*4.4. Figures*

## 5. Data analysis

*5.1. Census poverty data*

Data for this example comes from the U.S. Census Bureau's decennial census from 1960-2000, and from its American Community Survey in 2006. The variable being modeled is the logit of poverty rate by county in the states of Minnesota, Iowa, Wisconsin, Illinois, Indiana, and Michigan. A GWR model was fit to the data based on the predictors `pag`, `pex`, `pman`, `pserve`, `pfire`, `potprof`, `pwh`, `pblk`, `phisp`, and `metro`. One model was fit to the data from each census.

*5.2. Figures*

## 6. References


Antoniadas, A., I. Gijbels, and A. Verhasselt (2012). Variable selection in varying-coefficient models using p-splines. *Journal of Computational and Graphical Statistics 21*(3), 638–661.

Breiman, L. (1995). Better subset wregression using the nonnegative garrote. *Technometrics 51*, 373–384.

Efron, B. and R. Tibshirani (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science 1*(1), 54–75.

Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics 27*(5), 1491–1518.

Fotheringham, A., C. Brunsdon, and M. Charlton (2002). *Geographically weighted regression: the analysis of spatially varying relationships.* Wiley.

Hastie, T. and C. Loader (1993). Local regression: automatic kernel carpentry. *Statistical Science 8*(2), 120–143.

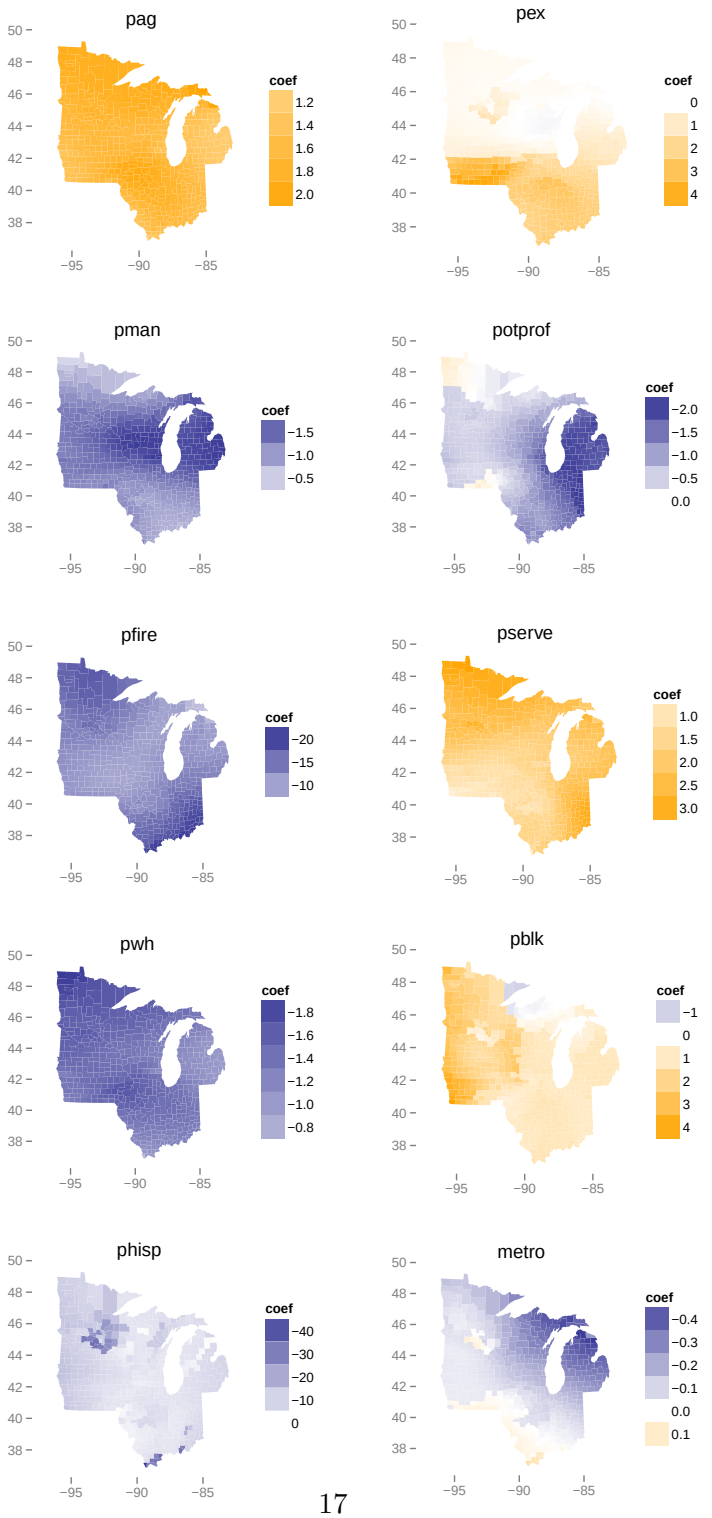





Figure 1: Estimated coefficient surfaces for the 1960 census.

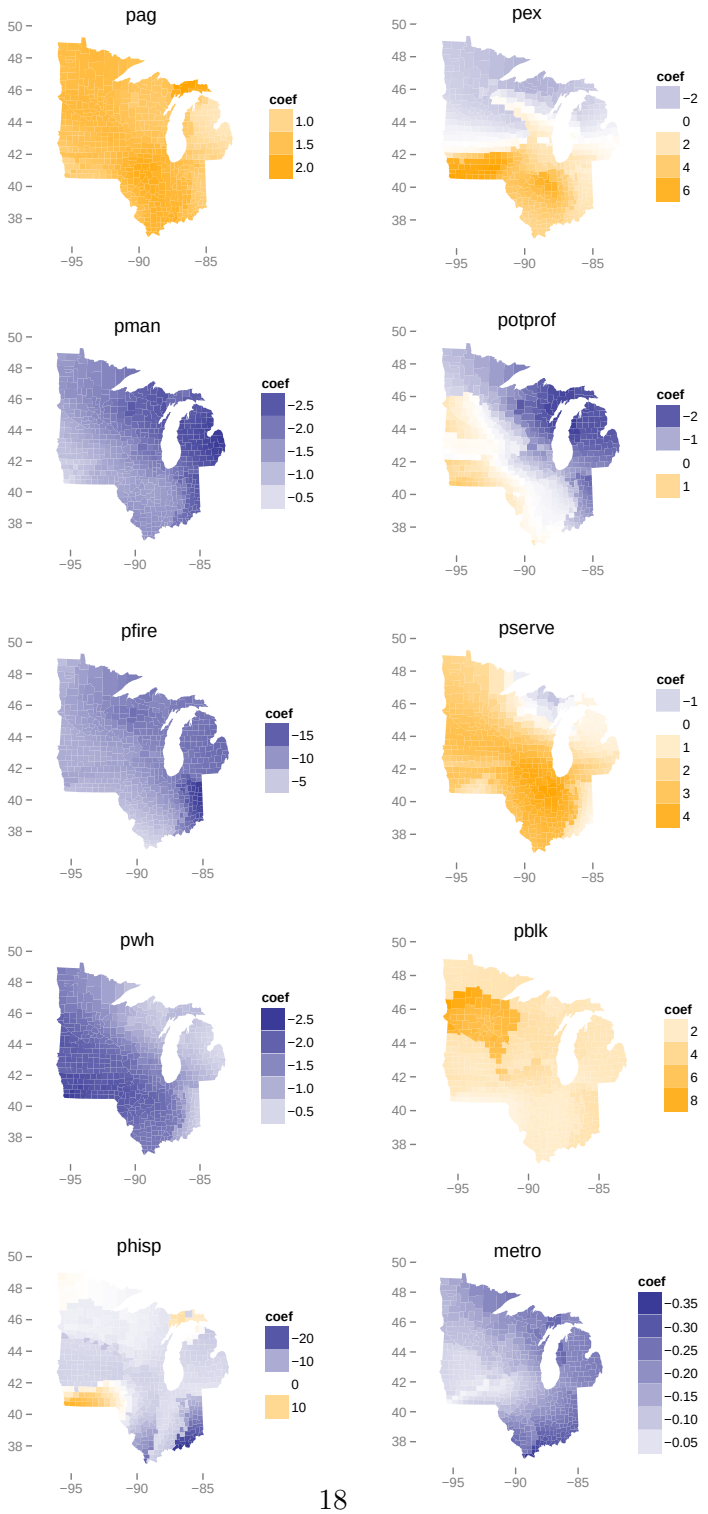





Figure 2: Estimated coefficient surfaces for the 1970 census.

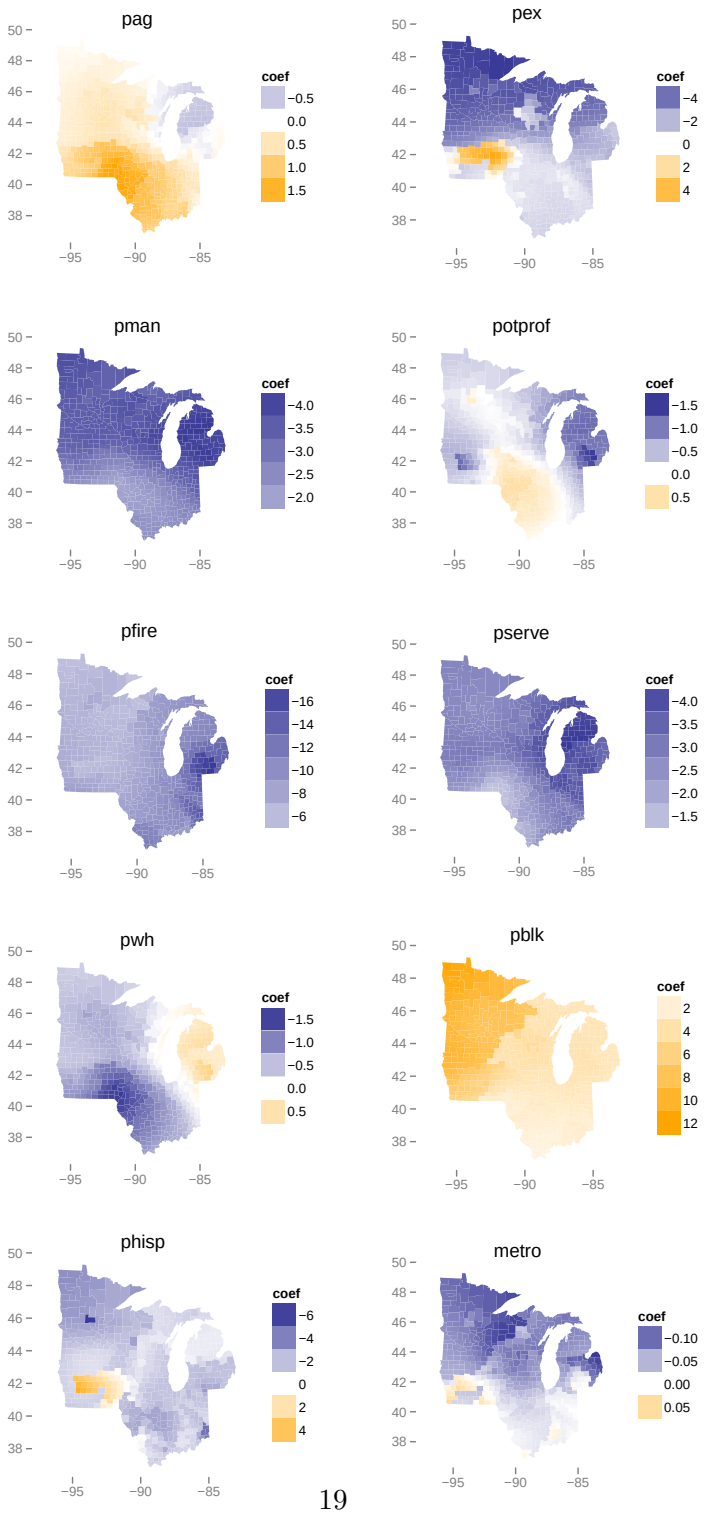





Figure 3: Estimated coefficient surfaces for the 1980 census.

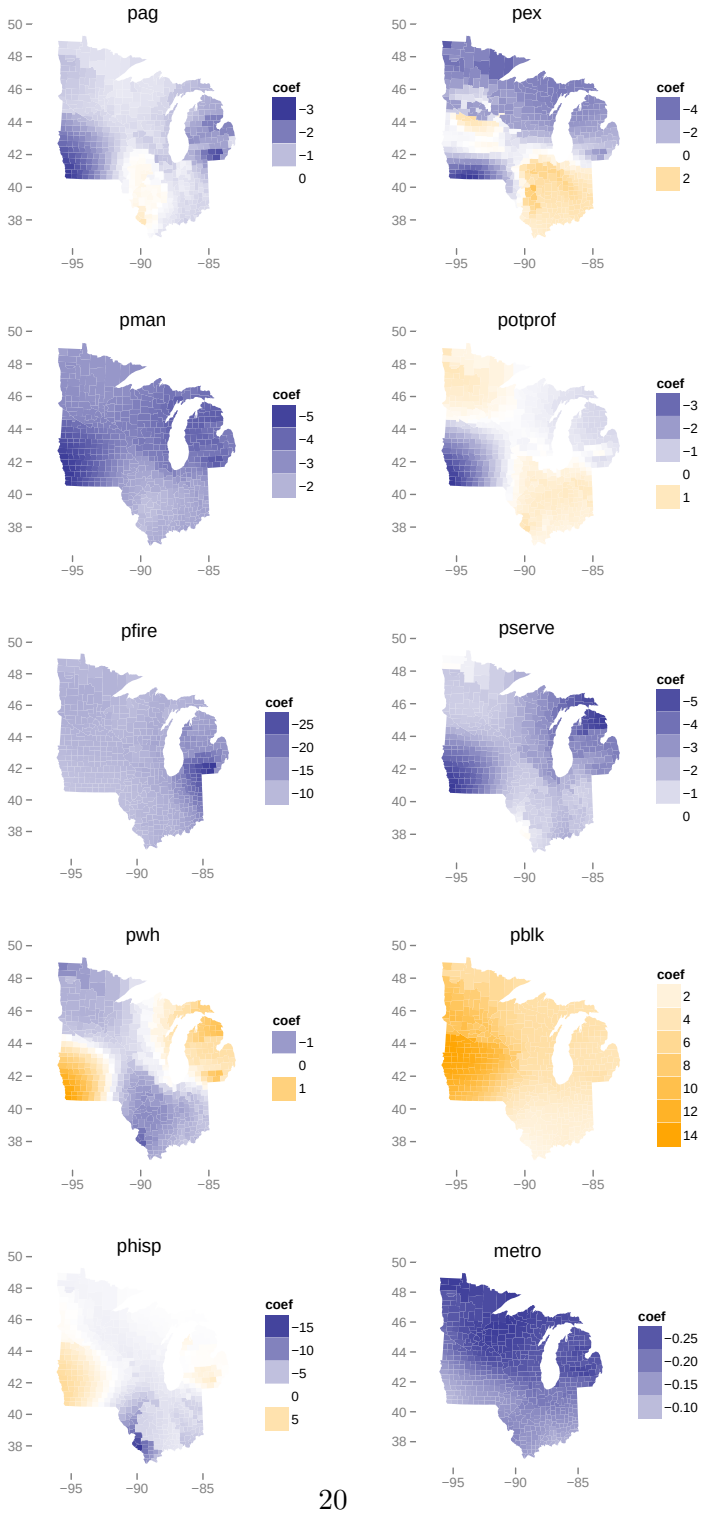





Figure 4: Estimated coefficient surfaces for the 1990 census.

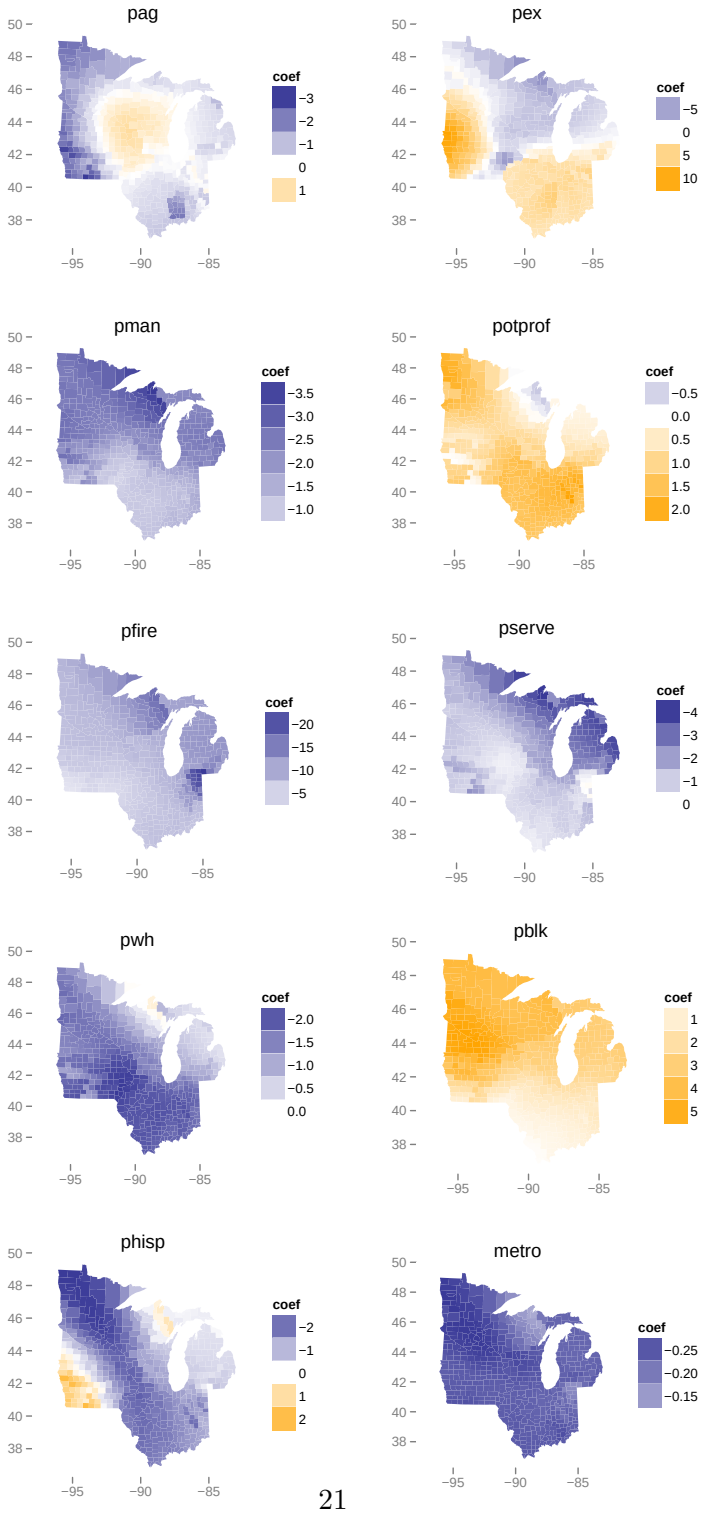





Figure 5: Estimated coefficient surfaces for the 2000 census.

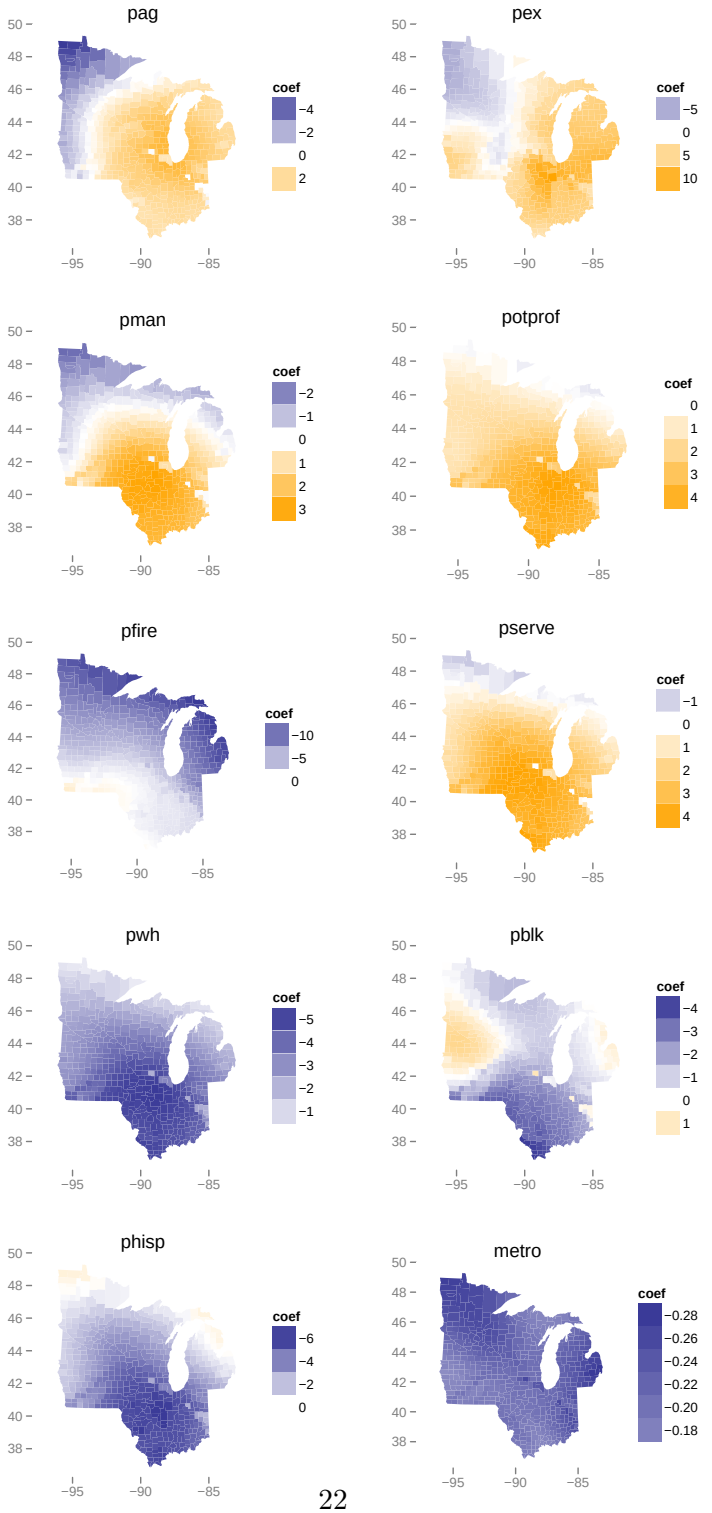





Figure 6: Estimated coefficient surfaces for the 2006 census.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological) 55*(4), pp. 757–796.

Loader, C. (1999). *Local regression and likelihood*. Springer New York.

Wang, L., H. Li, and J. Z. Huang (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association 103*(484), 1556–1569.

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.

Wheeler, D. C. (2009). Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and Planning A 41*, 722–742.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

Zou, H., T. Hastie, and R. Tibshirani (2007). On the "degrees of freedom" of the lasso. *Annals of Statistics 35*(5), 2173–2192.