

Local variable selection and parameter estimation for spatially varying coefficient models

Wesley Brooks

Department of Statistics
University of Wisconsin–Madison

December 12, 2013

These slides were prepared for a practice version of my preliminary exam to advance to Ph.D candidacy in statistics at the University of Wisconsin–Madison.

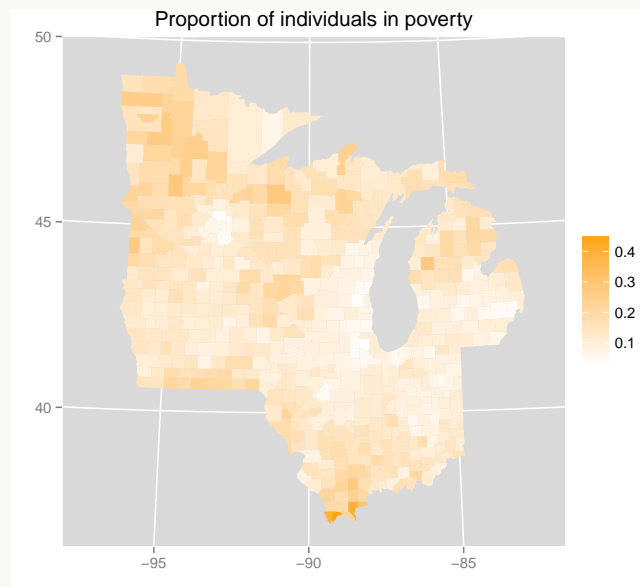
Motivation

2

New Section

Motivation

Response variable

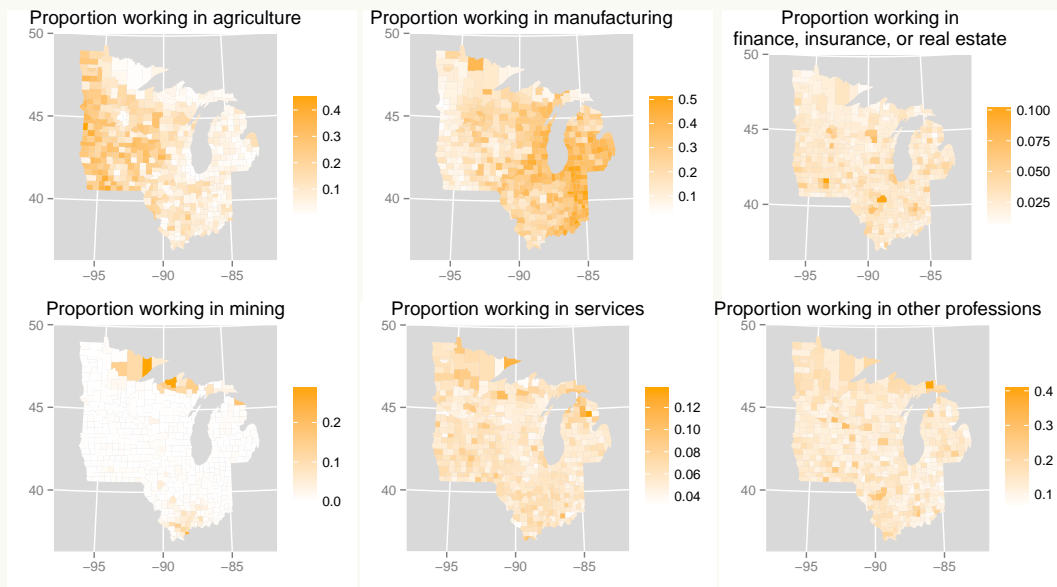


3

This is the county-level poverty rate from 1970

Motivation

Covariates



4

Here we have the proportion of people in each county who worked in manufacturing, agriculture, and services in 1970.

How is this data to be analyzed?

Motivation

Scientific questions

- ▶ Which of the economic-structure variables is associated with poverty rate?
- ▶ What are the sign and magnitude of that association?
- ▶ Is poverty rate associated with the same economic-structure variables across the entire region?
- ▶ How do the sign and magnitude of the associations vary across the region?

5

These are some sensible questions to ask about the county-level poverty rate. The work I'm presenting today attempts to answer these questions.

There are several other methods to answer at least some of these questions, which we'll cover next.

Introduction

6

New Section

Introduction

An overview

- ▶ Spatial regression
- ▶ Varying coefficient regression
 - Splines
 - Kernels
 - Wavelets
- ▶ Model selection via regularization

7

The existing methods to address the questions draw from these areas. Behind the methodology that I'm discussing is a wide range of literature.

Introduction

Definitions

- ▶ Univariate spatial response process $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$
- ▶ Multivariate spatial covariate process $\{\mathbf{X}(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$
- ▶ n = number of observations
- ▶ p = number of covariates
- ▶ Location (2-dimensional) \mathbf{s}
- ▶ Spatial domain \mathcal{D}

We'll use these variables throughout.

Introduction

Types of spatial data

► Geostatistical data:

- Observations are made at sampling locations s_i for $i = 1, \dots, n$
- E.g. elevation, temperature

► Areal data:

- Domain is partitioned into n regions $\{D_1, \dots, D_n\}$
- The regions do not overlap, and they divide the domain completely: $\mathcal{D} = \bigcup_{i=1}^n D_i$
- Sampling locations s_i for $i = 1, \dots, n$ are the centroids of the regions
- E.g. poverty rate, population, spatial mean temperature

The method I'm describing applies to geostatistical data, or to areal data when the observations are assumed to be located at the centroid.

The poverty data example is areal data; the simulation study I'll present later is based on simulated geostatistical data.

Introduction

Spatial linear regression (Cressie, 1993)

- ▶ A typical spatial linear regression model

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\boldsymbol{\beta} + W(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- ▶ $W(\mathbf{s})$ is a spatial random effect that accounts for autocorrelation in the response variable
- ▶ $\text{cov}(W(\mathbf{s}), W(\mathbf{t}))$: Matérn class
- ▶ The coefficients $\boldsymbol{\beta} = (1, \beta_1, \dots, \beta_p)$ are constant
- ▶ Relies on *a priori* global variable selection

10

Here we have the usual spatial regression as described by Cressie in his 1993 book.

This model assumes that the model coefficients are constant across the spatial domain and that the residuals can be separated into :

- The spatial random effect W that captures autocorrelation of the response, and - epsilon, which is iid white noise

The autocorrelation of the W 's is from a Matérn class covariance function, like the exponential covariance function.

This model relies on a priori model selection.

Typically Bayesian methods are used to estimate the coefficients.

Introduction

Spatially varying coefficient model (Gelfand et al., 2003)

- ▶ A more flexible model: coefficients in a spatial regression model can vary

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\boldsymbol{\beta}(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- ▶ $\{\beta_0(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}, \dots, \{\beta_p(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ are stationary spatial processes with Matérn covariance functions
- ▶ Still relies on *a priori* global variable selection

11

- The spatial random effect has been incorporated into the spatially varying intercept
- $\{\beta_1(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}, \dots, \{\beta_p(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ are stationary spatial processes with Matérn covariance functions
- Still relies on a priori global variable selection

The spatial regression model can be made more flexible by using stationary spatial processes to model the coefficients, rather than constants. The method was introduced by Gelfand in 2003.

As for the autocorrelated errors W in the traditional spatial regression, the coefficient processes have matern class covariance functions.

The autocorrelated errors W are now incorporated in the spatially varying intercept process.

This model also relies on a priori model selection and uses Bayesian methods to estimate the coefficients.

Introduction

Varying coefficients regression (VCR) (Hastie and Tibshirani, 1993)

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\boldsymbol{\beta}(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- ▶ Assume an effect modifying variable \mathbf{s}
- ▶ Coefficients are functions of \mathbf{s}

12

The varying coefficient regression model was described by Hastie and Tibshirani in 1993. This model looks like the spatially varying coefficient process, but is more general.

Any effect modifying variable \mathbf{s} - not just the spatial location - can be used in this model, and there are non-Bayesian methods to fit the model. We'll look at three.

Introduction

Spline-based VCR models (Wood, 2006)

- ▶ Splines are a way to parameterize smooth functions
- ▶ Estimate the varying coefficients via splines:

$$E\{Y(t)\} = \beta_1(t)X_1(t) + \cdots + \beta_p(t)X_p(t)$$

13

There is a good overview in Simon Wood's 2006 book of how to use regression splines to fit a varying coefficients regression model.

Regression splines are a way to parameterize a smooth function. In this case, the coefficient is a smooth function of the spatial location

fitting a spline-based VCR requires a priori model selection(?).

Introduction

Global selection in spline-based VCR models

Regularization methods for global variable selection in VCR models:

- ▶ The integral of a function squared (e.g. $\int \{f(t)\}^2 dt$) is zero if and only if the function is zero everywhere.
- ▶ Use regularization to encourage coefficient functions to be zero
 - SCAD penalty (Wang, Li, and Huang, 2008)
 - Non-negative garrote penalty (Antoniadas, Gijbels, and Verhasselt, 2012)

14

There are at least two references that describe how to select the covariates for a spline-based VCR model. Both rely on regularization.

The regularization penalizes the smooth coefficient function for being non-zero. Antoniadis et al. used a non-negative garrote penalty and Wang et al. used a SCAD penalty.

These selection methods are global - that is, they select variables for the entire domain simultaneously.

Introduction

Wavelet methods for VCR models

- ▶ Wavelet methods: decompose coefficient function into local frequency components
- ▶ Selection of nonzero local frequency components with nonzero coefficients:
 - Bayesian variable selection (Shang, 2011)
 - Lasso (J. Zhang and Clayton, 2011)
- ▶ Sparsity in the local frequency components; not in the local covariates

15

Another way to fit a VCR model is to use a wavelet decomposition, which decomposes the coefficient function into its local frequency components. Model selection is then used to identify which local frequency components to use in the model.

Murray Clayton's students Zuofeng Shang and Jun Zhang used Bayesian variable selection and the Lasso, respectively, to select the local frequency components.

However, these methods achieve sparsity in the wavelet coefficients, which does not imply sparsity in the covariates. So these methods don't achieve local model selection.

Now let's take a look at geographically weighted regression.

Geographically weighted regression

16

New Section

Geographically weighted regression

(Brundson, S. Fotheringham, and Martin Charlton, 1998;
A. Fotheringham, Brunsdon, and M. Charlton, 2002)

- ▶ Consider observations at sampling locations s_1, \dots, s_n
- ▶ $y(s_i) = y_i$ the univariate response at location s_i
- ▶ $x(s_i) = x_i$ the $(p + 1)$ -variate vector of covariates at location s_i
- ▶ Assume $y_i = x_i' \beta_i + \varepsilon_i$ where $\varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$

17

Geographically weighted regression is the method of using local regression to estimate the coefficients in a spatially varying coefficient regression model.

Our sampling locations are called s , the response is y and the covariates (which number p) are called x .

Assume that the errors are iid normal.

Geographically weighted regression

(Brundson, S. Fotheringham, and Martin Charlton, 1998;
A. Fotheringham, Brunsdon, and M. Charlton, 2002)

- ▶ The total log likelihood is
$$\ell(\boldsymbol{\beta}) = - (1/2) \left\{ n \log(2\pi\sigma^2) + \sigma^{-2} \sum_{i=1}^n (y_i - \mathbf{x}_i' \boldsymbol{\beta}_i)^2 \right\}$$
- ▶ With n observations and $n(p+1)$ parameters, the model is not identifiable.
- ▶ Idea: to estimate parameters by borrowing strength from nearby observations

18

We have here the total log likelihood of the observed data.

since there are n observations and $n \times (p+1)$ free parameters, the model is not identifiable.

We will estimate the parameters by borrowing strength from nearby observations

Geographically weighted regression

Local regression (Loader, 1999)

Local regression uses a kernel function at each sampling location to weight observations based on their distance from the sampling location.

$$\mathcal{L}_i = \prod_{i'=1}^n (\mathcal{L}_{i'})^{w_{ii'}}$$
$$\ell_i = \sum_{i'=1}^n w_{ii'} \left\{ \log(\sigma_i^2) + \sigma_i^{-2} (y_{i'} - \mathbf{x}_{i'}' \boldsymbol{\beta}_i)^2 \right\}$$

Given the weights, a local model is fit at each sampling location using the local likelihood

19

Local regression uses a kernel function at each sampling location to weight the observations. For a GWR model, the kernel weights are based on an observation's distance from the sampling location.

Here we have the likelihood at one sampling location. Note that each observation is given a weight $w(i, i')$

Given the weights, a local model is fit at each sampling location using the local likelihood

Maximizing the local likelihood for a model of Gaussian data with iid errors can be done by weighted least squares.

Geographically weighted regression

Local likelihood (Loader, 1999)

Weights are calculated via a kernel, e.g. the bisquare kernel:

$$w_{ii'} = \begin{cases} \left\{ 1 - (\phi^{-1} \delta_{ii'})^2 \right\}^2 & \text{if } \delta_{ii'} < \phi, \\ 0 & \text{if } \delta_{ii'} \geq \phi \end{cases} \quad (1)$$

where

- ▶ ϕ is a bandwidth parameter
- ▶ $\delta_{ii'} = \delta(\mathbf{s}_i, \mathbf{s}_{i'}) = \|\mathbf{s}_i - \mathbf{s}_{i'}\|_2$ is the Euclidean distance between sampling locations \mathbf{s}_i and $\mathbf{s}_{i'}$.

20

This is the form of the bisquare kernel, which is what I've used in this work.

ϕ is a bandwidth parameter and $\delta(i, i')$ is the distance between points i and i' .

Geographically weighted regression

Bandwidth estimation via the AIC_c (Hurvich, Simonoff, and Tsai, 1998)

- ▶ Smaller bandwidth: less bias, more flexible coefficient surface
- ▶ Large bandwidth: less variance, less flexible coefficient surface
- ▶ Idea: to estimate the degrees of freedom used in estimating the coefficient surface:
 - $\hat{y} = Hy$
 - $\nu = \text{tr}(H)$
- ▶ Then the corrected AIC for bandwidth selection is:

$$AIC_c = 2n \log \sigma + n \left\{ \frac{n + \nu}{n - 2 - \nu} \right\}$$

21

Selecting the bandwidth in a GWR model is done by the AIC. There is a bias-variance tradeoff in selecting an optimal bandwidth.

Local variable selection and parameter estimation

22

New Section

Geographically weighted regression

Geographically weighted Lasso (Wheeler, 2009)

Within a GWR model, using the Lasso for local variable selection is called the geographically weighted Lasso (GWL).

- ▶ The GWL requires estimating a Lasso tuning parameter for each local model
- ▶ (Wheeler, 2009) estimates the local Lasso tuning parameter at location s_i by minimizing a jackknife criterion: $|y_i - \hat{y}_i|$
- ▶ The jackknife criterion can only be calculated where data are observed, making it impossible to use the GWL to impute missing data or to estimate the value of the coefficient surface at new locations
- ▶ Also, the Lasso is known to be biased in variable selection and suboptimal for coefficient estimation

23

For local model selection in a GWR model, Wheeler proposed the geographically weighted lasso (GWL) in 2009.

At each model location, the Lasso is used to select the locally-relevant predictors

The GWL uses a jackknife criterion to select the local lasso tuning parameters, which means the GWL cannot be used at model locations other than sample locations.

That means the GWL cannot be used for interpolating the coefficient surface or for imputing missing values of the response variable.

Local variable selection and parameter estimation

Geographically weighted adaptive Lasso (GWAL)

- ▶ Local variable selection in a GWR model using the adaptive Lasso (AL) (Zou, 2006)
- ▶ Under suitable conditions, the AL has an oracle property for selection
- ▶ Let

$$\begin{aligned}\mathcal{S}(\beta_i) &= -2\ell_i(\beta_i) + \mathcal{J}_2(\beta_i) \\ &= \sum_{i'=1}^n w_{ii'} \left\{ \log(\sigma_i^2) + \sigma_i^{-2} (y_{i'} - \mathbf{x}_{i'}' \beta_i)^2 \right\} \\ &\quad + \lambda_i \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij}\end{aligned}$$

24

The geographically weighted adaptive lasso (GWAL) overcomes these shortcomings of the GWL.

The GWAL uses the adaptive lasso, which has an oracle property under suitable conditions.

Here is the penalized likelihood for a local GWAL model

The adaptive lasso consists of applying a different penalty to each covariate based on the adaptive weights, which are derived from the weighted least squares coefficients.

The adaptive weights $\gamma_i = (\gamma_{i1}, \dots, \gamma_{ip})'$ are defined in the same way as for the AL, and the elastic net parameter $\alpha_i \in [0, 1]$ controls the balance between ℓ_1 penalty $\lambda_i \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij}$ and ℓ_2 penalty $\lambda_i \sum_{j=1}^p (\beta_{ij} / \gamma_{ij})^2$.

Local variable selection and parameter estimation

Geographically weighted adaptive elastic net (GWAL)

Note:

- ▶ $\sum_{i'=1}^n w_{ii'} (y_{i'} - \mathbf{x}_{i'}' \boldsymbol{\beta}_i)^2$ is the weighted sum of squares minimized by traditional GWR
- ▶ $\mathcal{J}_1(\boldsymbol{\beta}_i) = \lambda_i \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij}$ is the AL penalty.

25

The adaptive weights $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{ip})'$ are defined in the same way as for the AL, and the elastic net parameter $\alpha_i \in [0, 1]$ controls the balance between ℓ_1 penalty $\lambda_i^* \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij}$ and ℓ_2 penalty $\lambda_i^* \sum_{j=1}^p (\beta_{ij} / \gamma_{ij})^2$.

Local variable selection and parameter estimation

Tuning parameter estimation

To estimate an AL tuning parameter for each local, use a local BIC that allows fitting a local model at any location within the spatial domain

$$\begin{aligned}\text{BIC}_i &= -2 \sum_{i'=1}^n \ell_{ii'} + \log \left(\sum_{i'=1}^n w_{ii'} \right) \text{df}_i \\ &= \sum_{i'=1}^n w_{ii'} \left\{ \log(2\pi) + \log(\hat{\sigma}_i^2) + \hat{\sigma}_i^{-2} \left(y_{i'} - \mathbf{x}'_{i'} \hat{\boldsymbol{\beta}}_{i'} \right)^2 \right\} \\ &\quad + \log \left(\sum_{i'=1}^n w_{ii'} \right) \text{df}_i\end{aligned}$$

26

The adaptive weights $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{ip})'$ are defined in the same way as for the AL, and the elastic net parameter $\alpha_i \in [0, 1]$ controls the balance between ℓ_1 penalty $\lambda_i^* \sum_{j=1}^p |\beta_{ij}|/\gamma_{ij}$ and ℓ_2 penalty $\lambda_i^* \sum_{j=1}^p (\beta_{ij}/\gamma_{ij})^2$.

We treat the sum of the weights around the sampling location as the number of observations for the local BIC.

Local variable selection and parameter estimation

Geographically weighted adaptive elastic net (GWEN)

- Local variable selection in a GWR model using the adaptive elastic net (AEN) (Zou and H. Zhang, 2009)
- Under suitable conditions, the AEN has an oracle property for selection

$$\begin{aligned}\mathcal{S}(\beta_i) &= -2\ell_i(\beta_i) + \mathcal{J}_2(\beta_i) \\ &= \sum_{i'=1}^n w_{ii'} \left\{ \log \sigma_i^2 + (\sigma_i^2)^{-1} (y_{i'} - \mathbf{x}_{i'}' \beta_i)^2 \right\} \\ &\quad + \alpha_i \lambda_i^* \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij} \\ &\quad + (1 - \alpha_i) \lambda_i^* \sum_{j=1}^p (\beta_{ij} / \gamma_{ij})^2\end{aligned}$$

27

The geographically weighted adaptive elastic net (GWEN) is similar to the GWAL but uses the elastic net for local model selection

The adaptive elastic net also has an oracle property under suitable conditions.

The adaptive elastic net consists of adding an L2 penalty to the regularization in addition to the L1 penalty of the adaptive lasso.

Here is the penalized likelihood for a local GWEN model

The adaptive weights $\gamma_i = (\gamma_{i1}, \dots, \gamma_{ip})'$ are defined in the same way as for the AL, and the elastic net parameter $\alpha_i \in [0, 1]$ controls the balance between ℓ_1 penalty $\lambda_i^* \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij}$ and ℓ_2 penalty $\lambda_i^* \sum_{j=1}^p (\beta_{ij} / \gamma_{ij})^2$.

Local variable selection and parameter estimation

Geographically weighted adaptive elastic net (GWEN)

- The AEN penalty function is

$$\mathcal{J}_2(\boldsymbol{\beta}_i) = \alpha_i \lambda_i^* \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij} + (1 - \alpha_i) \lambda_i^* \sum_{j=1}^p (\beta_{ij} / \gamma_{ij})^2$$

28

The adaptive weights $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{ip})'$ are defined in the same way as for the AL, and the elastic net parameter $\alpha_i \in [0, 1]$ controls the balance between ℓ_1 penalty $\lambda_i^* \sum_{j=1}^p |\beta_{ij}| / \gamma_{ij}$ and ℓ_2 penalty $\lambda_i^* \sum_{j=1}^p (\beta_{ij} / \gamma_{ij})^2$.

Local variable selection and parameter estimation

Bandwidth parameter estimation

- ▶ Traditional GWR:
 - $\hat{y} = Hy$
 - So traditional GWR is a linear smoother
 - $\nu = \text{tr}(H)$ is the degrees of freedom for the model
- ▶ GWAL:
 - $\hat{y} = H^\dagger y + T^\dagger \gamma$
- ▶ GWEN:
 - $\hat{y} = H^* y + T^* \gamma$
- ▶ Neither GWEN nor GWAL is a linear smoother
 - df not equal to trace of projection matrix for GWAL, GWEN
- ▶ Solution: use GWEN or GWAL for selection then fit local model for the selected variables via traditional GWR
 - Now $\text{df} = \nu = \text{tr}(H)$

29

note

Local variable selection and parameter estimation

Locally linear coefficient estimation

- ▶ GWR, GWEN, GWAL: coefficients locally constant
 - as in Nadaraya-Watson kernel smoother
 - Leads to bias where there is a gradient at the boundary
- ▶ Solution: local polynomial modeling
 - First-order polynomial: locally linear coefficients
- ▶ Augment with covariate-by-location interactions
 - Two-dimensional
 - Augment with selected covariates only

30

note

Simulation study

31

New Section

Simulation study

Simulating covariates

- ▶ 30×30 grid on $[0, 1] \times [0, 1]$
- ▶ Five covariates $\tilde{X}_1, \dots, \tilde{X}_5$
- ▶ Gaussian random fields:

$$\begin{aligned}\tilde{X}_j &\sim N(0, \Sigma) \text{ for } j = 1, \dots, 5 \\ \{\Sigma\}_{i,i'} &= \exp\{-\tau^{-1} \delta_{ii'}\} \text{ for } i, i' = 1, \dots, n\end{aligned}$$

- ▶ Colinearity: ρ

note

Simulation study

Simulating the response

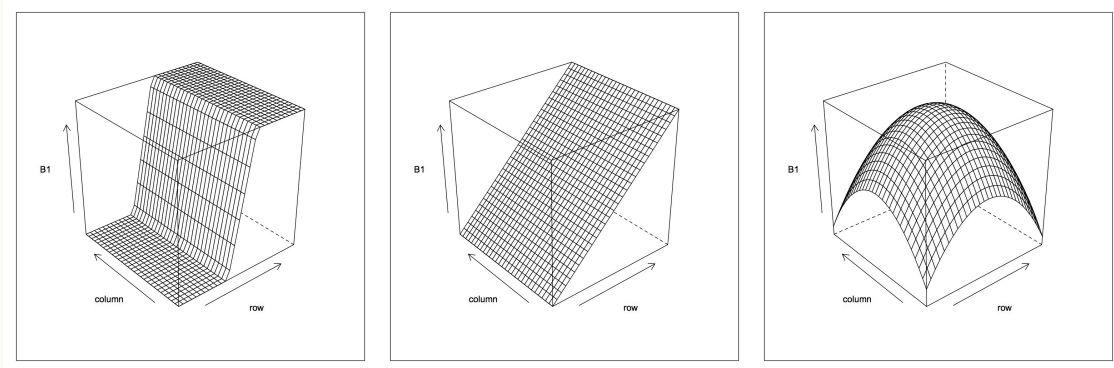
- ▶ $Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\boldsymbol{\beta}(\mathbf{s}) = \sum_{j=1}^5 \beta_j(\mathbf{s})X_j(\mathbf{s}) + \varepsilon(\mathbf{s})$
- ▶ $\varepsilon(\mathbf{s}) \sim iid\ N(0, \sigma^2)$
- ▶ $\beta_1(\mathbf{s})$, the coefficient function for X_1 , is nonzero in part of the domain.
- ▶ Coefficients for X_2, \dots, X_5 are zero everywhere

33

note

Simulation study

Coefficient functions: step, gradient, and parabola



34

note

Simulation study

Simulation settings

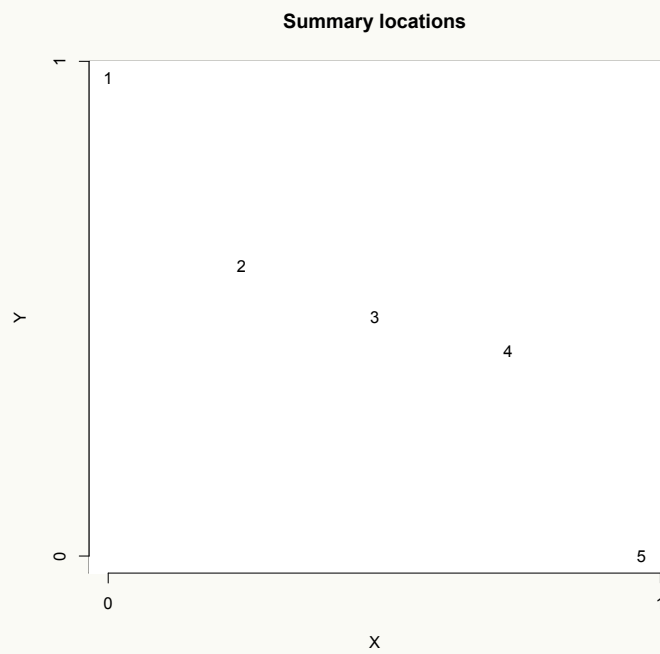
Each setting simulated 100 times:

Setting	function	ρ	σ^2
1	step	0	0.25
2	step	0	1
3	step	0.5	0.25
4	step	0.5	1
5	gradient	0	0.25
6	gradient	0	1
7	gradient	0.5	0.25
8	gradient	0.5	1
9	parabola	0	0.25
10	parabola	0	1
11	parabola	0.5	0.25
12	parabola	0.5	1

note

Simulation results

Summary locations



36

note

Simulation results

Selection performance

- ▶ Non-ambiguous locations (80):
 - 52 saw no false negatives
 - 72 had no false positives
 - 26 neither false positives nor false negatives
- ▶ Noise variance had more effect on selection than colinearity
- ▶ No difference between GWEN, GWAL

37

note

Simulation results

Estimation performance

- ▶ Oracular selection
 - best $\text{MSE}(\hat{\beta}_1)$ in 41 of the 60 cases
- ▶ Oracular selection
 - best $\text{MSE}(\hat{\beta}_1)$ in 41 of the 60 cases
- ▶ Generally small difference between GWR, oracular, GWEN-LLE, and GWAL-LLE
- ▶ Noise variance had more effect on selection than colinearity
- ▶ No difference between GWEN, GWAL
- ▶ Fitting \hat{y} : best MSE split evenly between

38

note

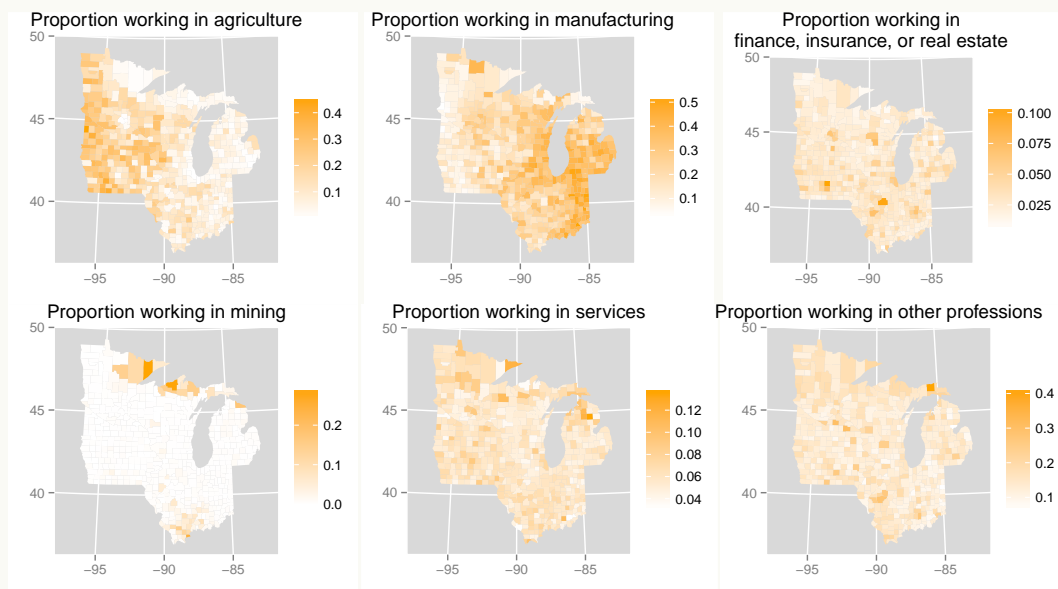
Data example: poverty rate in the upper
midwest

39

New Section

Data example: poverty rate in the upper midwest

Revisiting the introductory example



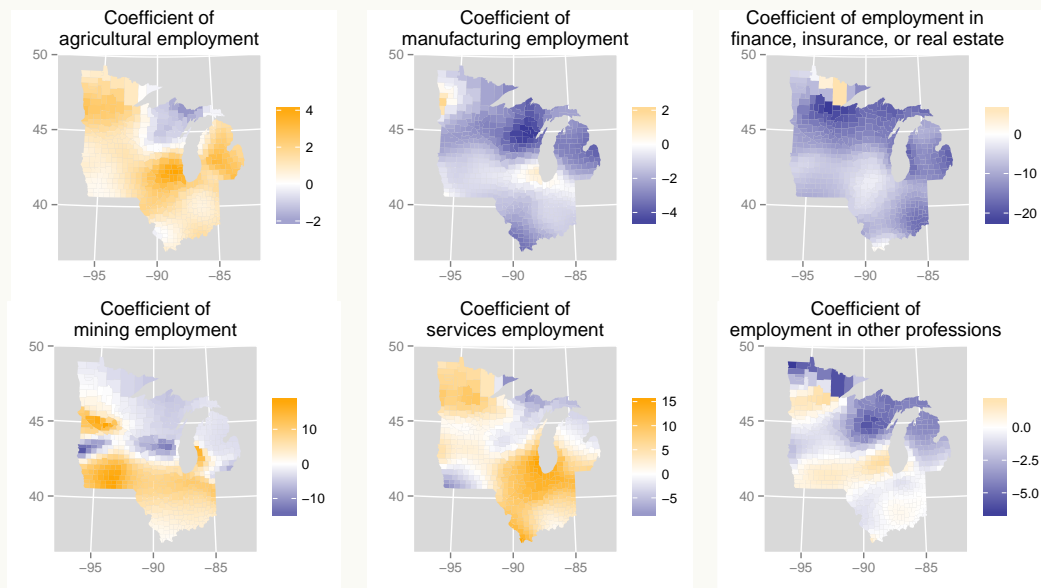
40

This is the county-level poverty rate from 1970, as well as the proportion of people who worked in manufacturing, agriculture, and services.

How is this data to be analyzed?

Data example: poverty rate in the upper midwest

Results from traditional GWR



note

Data example: poverty rate in the upper midwest

Data description

- ▶ Response: logit-transformed poverty rate in the Upper Midwest states of the U.S.
 - Minnesota, Iowa, Wisconsin, Illinois, Indiana, Michigan
- ▶ Covariates: employment structure (raw proportion employed in:)
 - agriculture
 - finance, insurance, and real estate
 - manufacturing
 - mining
 - services
 - other professions
- ▶ Data source: U.S. Census Bureau's decennial census of 1970

42

note

Data example: poverty rate in the upper midwest

Data description

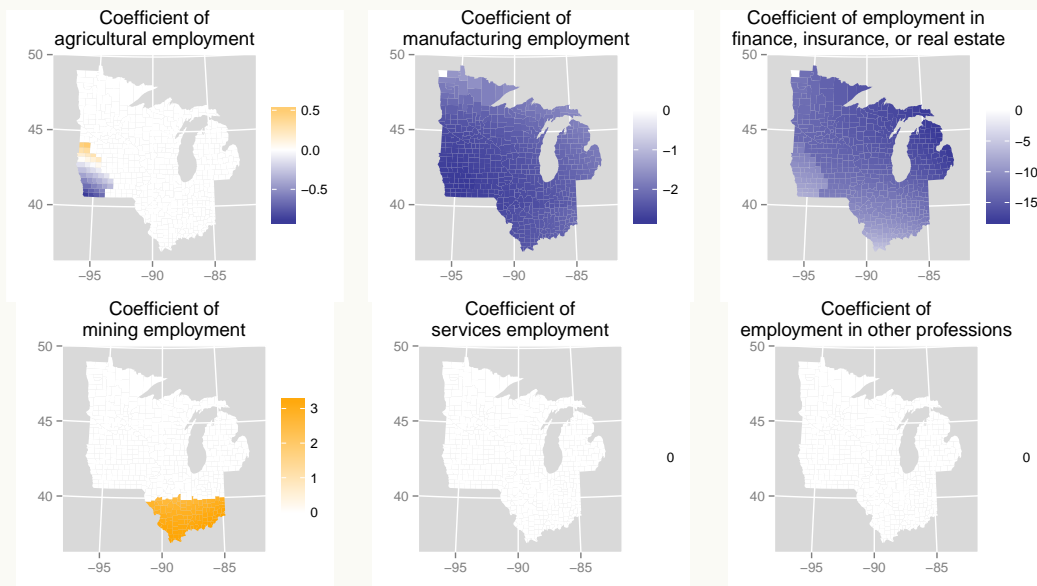
- ▶ Data aggregated to the county level
 - counties are areal units
- ▶ county centroid treated as sampling location

43

note

Data example: poverty rate in the upper midwest

Results from GWEN



note

Data example: poverty rate in the upper midwest

Results from GWEN-LLE

- ▶ Relatively constant compared to GWR
- ▶ Services, "other professions" do not affect the poverty rate
- ▶ Manufacturing: negative coefficient everywhere
- ▶ Finance, insurance, and real estate negative coefficient everywhere
 - Largest magnitude (min: -20, next-largest: -3)
 - GWR comparable to GWEN-LLE
- ▶ Manufacturing: negative coefficient everywhere
 - GWR: coefficient greater than zero near Chicago and in NW Minnesota
- ▶ Agriculture: nonzero in western Iowa
 - North-south gradient to coefficient
 - ranges positive to negative
- ▶ Mining: nonzero in parts south
 - Associated with increased poverty rate
 - Comparable to GWR within far southern range

45

note

Future work

46

New Section

Future work

- ▶ Apply the GWEN to models for non-Gaussian response variable
- ▶ Incorporate spatial autocorrelation in the model (simulated errors were iid)

note

Acknowledgements