

# Local variable selection and parameter estimation for spatially varying coefficient models

Wesley Brooks

Department of Statistics  
University of Wisconsin–Madison

January 17, 2014

Hello, I'm Wesley Brooks. My undergraduate work was in electrical engineering and math at the University of Alaska, and I completed my M.S. in this department in 2012. Today I'm presenting work toward my Ph.D.

The subject of this research is local variable selection and parameter estimation for spatially varying coefficient regression models.

I'll begin by discussing an example problem that helps to motivate this work.

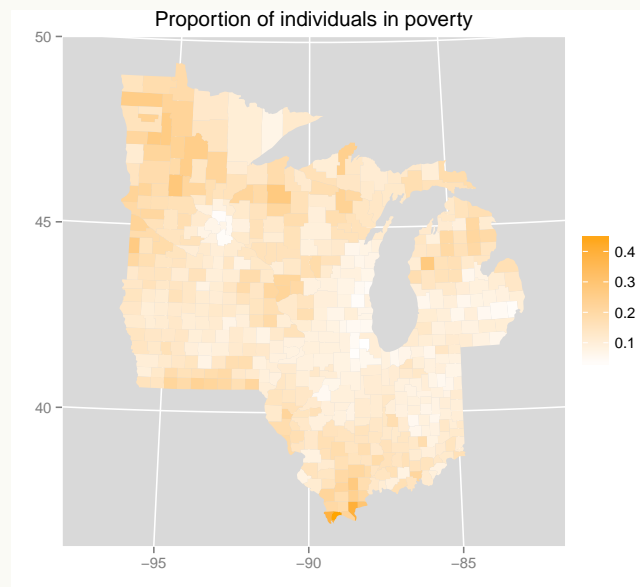
# Motivation

2

New Section

# Motivation

Response variable



3

This is the county-level poverty rate in the upper midwestern states from the US census in 1970

and...

# Motivation

## Covariates



4

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

A researcher may approach this data wanting to understand how the poverty rate is related to these variables that describe the economic structure.

# 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 scientific questions to ask about the relationship between economic structure and the county-level poverty rate.

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?

The work I'm presenting today attempts to answer these questions and focuses on selecting variables and estimating coefficients locally. I'll begin by discussing some of the methods that can currently be used to approach questions of this kind.

# Introduction

6

New Section

# Introduction

An overview

- ▶ Definitions
- ▶ Spatial regression
- ▶ Varying coefficient regression
  - Splines
  - 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.

Spatial regression

Varying coefficient regression

Splines

Wavelets

Model selection via regularization

# 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  $\mathcal{D}$

8

For the remainder of this talk, I'll be using these terms.

The response  $Y$  is a univariate spatial process.

$\mathbf{X}$  is a multivariate spatial process of covariates.

$n$  is the number of observations,  $p$  is the number of covariates,  $\mathbf{s}$  indicates the two dimensional spatial location, and indexes that spatial domain  $\mathcal{D}$ .



# 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
- ▶  $\varepsilon(\mathbf{s})$  is iid random noise
- ▶ The coefficients  $\boldsymbol{\beta} = (1, \beta_1, \dots, \beta_p)$  are constant
- ▶ Requires *a priori* global variable selection

9

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

This model extends the traditional linear regression by separating the residuals into:

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

The spatial regression model estimates the coefficients as constants, and the model relies on a priori global variable selection.

# 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
- ▶ Requires *a priori* global variable selection

10

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

The random effect  $W$  from the previous slide is now incorporated in the spatially varying intercept process.

This model also relies on *a priori* global variable selection.

# 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  $\mathbf{s}$

11

The varying coefficient regression model was described by Hastie and Tibshirani in 1993. The form of this model looks like the spatially varying coefficient process, but this model is more general because the coefficients are not necessarily spatial processes in this model.

In fact, the effect-modifying variable  $\mathbf{s}$  does not necessarily need to represent spatial location.

There are many ways to fit a varying coefficient regression model. Splines and kernel-based local regression are the most common.

# Introduction

Global selection in spline-based VCR models

Global variable selection can be done via regularization in a VCR model where the coefficient functions are modeled as splines

- ▶ The  $L_2$  norm of a function (e.g.  $\int \{f(t)\}^2 dt$ ) is zero if and only if the function is zero everywhere.
- ▶ Use regularization to penalize nonzero coefficient functions
  - SCAD penalty and B-splines (Wang *et al.*, 2008a)
  - Non-negative garrote penalty and P-splines (Antoniadis *et al.*, 2012b)

12

Splines are a method of parameterizing smooth functions as a sum of smooth basis functions

It is possible to use splines to represent the coefficients in a varying coefficient regression model and there is a good overview of this topic in Simon Wood's 2006 book.

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

The regularization penalizes the smooth coefficient function for being non-zero. Antoniadis *et al.* used P-splines with a non-negative garrote penalty on the  $L_2$  norm and Wang *et al.* used B-splines with a SCAD penalty on the  $L_2$  norm.

These methods select variables globally - that is, the variables are included or excluded on the entire domain. That doesn't suit our goal of selecting locally relevant covariates.

# 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 (Zhang and Clayton, 2011)
- ▶ Sparsity in the local frequency components; not in the local covariates

13

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 relevant 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 variable selection.

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

# Geographically weighted regression

14

New Section

# Geographically weighted regression

Brundson *et al.* (1998), Fotheringham *et al.* (2002)

- ▶ Consider observations at sampling locations  $\mathbf{s}_1, \dots, \mathbf{s}_n$
- ▶  $y(\mathbf{s}_i) = y_i$  the univariate response at location  $\mathbf{s}_i$
- ▶  $\mathbf{x}(\mathbf{s}_i) = \mathbf{x}_i$  the  $p$ -vector of covariates at location  $\mathbf{s}_i$
- ▶ Assume  $y_i = \mathbf{x}_i' \boldsymbol{\beta}_i + \varepsilon_i$  where  $\varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$

15

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  $\mathbf{s}$ , the response is  $y$  and the covariates (which number  $p$ ) are called  $\mathbf{x}$ .

Assume that the errors are iid normal.

The notation  $\boldsymbol{\beta}_i$  is used to indicate that the coefficients are specific to location  $i$ .

# Geographically weighted regression

Brundson *et al.* (1998), Fotheringham *et al.* (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  $np + 1$  parameters, the model is not identifiable.
- Idea: to estimate parameters by borrowing strength from nearby observations

16

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

Because each  $\boldsymbol{\beta}_i$  is a  $p$ -vector of local coefficients, this model has  $n$  observations and  $np+1$  parameters, so 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.

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

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

17

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  $w_{ii'}$

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

- ▶  $\phi$  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  $\mathbf{s}_{i'}$ .

18

The local weights  $w_{ii'}$  from the previous slide are calculated from a kernel.

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

$\phi$  is a bandwidth parameter and  $\delta_{ii'}$  is the distance between points  $i$  and  $i'$ .

# Geographically weighted regression

Bandwidth estimation via the  $AIC_c$  (Hurvich *et al.*, 1998)

- ▶ Smaller bandwidth: less bias, more flexible coefficient surface
- ▶ Large bandwidth: less variance, less flexible coefficient surface
- ▶ Choose the bandwidth parameter to optimize the bias-variance tradeoff

19

To estimate a GWR model, it is necessary to estimate the bandwidth parameter, which involves a bias-variance tradeoff.

When the bandwidth is small, the coefficient surface is flexible and should have less bias but greater variance.

When the bandwidth is large, the coefficient surface is less flexible so it has less variance but potentially more bias.

# Geographically weighted regression

Bandwidth estimation via the  $AIC_c$  (Hurvich *et al.*, 1998)

- The corrected AIC for bandwidth selection is:

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

- $\hat{y} = H y$
- $\nu = \text{tr}(H)$
- $H_j = \{ W X (X' W X)^{-1} X \}_j$
- Where subscript  $j$  indicates the  $j$ th row of the matrix

20

One way to estimate the GWR bandwidth is via the corrected AIC of Hurvich et al..

## Local variable selection and parameter estimation

21

New Section

# Geographically weighted Lasso

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:  $|y_i - \hat{y}_i^{(-i)}|$
- ▶ 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
- ▶ Lasso not generally unbiased in variable selection (Fan and Li, 2001; Zou, 2006)

22

For local variable 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.

Furthermore, the Lasso is generally not unbiased in variable selection, as was suggested in the Fan and Li 2001 paper and proven in Hui Zou's 2006 paper.

## Local variable selection and parameter estimation

Geographically weighted adaptive elastic net (GWEN)

To overcome the shortcomings of the GWL, use a variable selection technique with potential for oracle properties and a tuning technique that can be applied anywhere on the domain.

- ▶ Local variable selection via the adaptive elastic net (AEN) (Zou and Zhang, 2009)
- ▶ Tuning parameters for local models selected by the BIC (Schwarz, 1978)

23

A modeling method with the potential for an oracle property in variable selection and that can be tuned at model locations other than sampling locations would be superior to the GWL.

I'm proposing in this work the geographically weighted adaptive elastic net, or the GWEN for short. It is a technique for fitting spatially varying coefficient models that uses the adaptive elastic net for local variable selection and the BIC to select the tuning parameters for the local models.

# Local variable selection and parameter estimation

Geographically weighted adaptive elastic net (GWEN)

The adaptive elastic net:

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

Where

$$\gamma_i = (\mathbf{X}' \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}_i \mathbf{Y}$$

24

The adaptive elastic net is a penalized regression technique that applies a different penalty to each covariate based on adaptive weights, which are derived from traditional GWR's weighted least squares coefficients.

S here is the penalized likelihood for a local GWEN model. The elastic net incorporates both an  $\ell_1$  and an  $\ell_2$  penalty

The adaptive weights  $\gamma_i$  are defined in the same way as for the AL, and

the elastic net parameter  $\alpha_i \in [0, 1]$  controls the balance between  $\ell_1$  penalty and  $\ell_2$  penalty



# Local variable selection and parameter estimation

Geographically weighted adaptive elastic net (GWAL)

- ▶ The adaptive Lasso penalty function is (Zou, 2006)

$$\mathcal{J}_1(\beta_i) = \lambda_i \sum_{j=1}^p |\beta_{ij}| / |\gamma_{ij}|$$

- ▶ Under suitable conditions, the AL has an oracle property for selection in linear regression

25

The geographically weighted adaptive lasso (GWAL) is a particular case of the GWEN that uses the adaptive lasso for variable selection.

The adaptive Lasso penalty consists of an L1 penalty with adaptive weights.

The adaptive lasso also has an oracle property under suitable conditions in linear regression.

# Local variable selection and parameter estimation

Tuning parameter estimation

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

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

26

Variable selection via the adaptive elastic net requires selecting the tuning parameter  $\lambda$ . For the GWEN,  $\lambda$  is selected by the BIC.

We treat the sum of the weights around the sampling location as the number of observations for the local BIC. The number of variables selected for the model is the degrees of freedom of the local model.

This BIC can be computed for a model at any location within the domain, allowing the GWEN to be used for imputing missing data or interpolating a model between observation locations.

# Local variable selection and parameter estimation

## Bandwidth parameter estimation

- ▶ Traditional GWR:
  - $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$
  - So traditional GWR is a linear smoother
  - $\nu = \text{tr}(\mathbf{H})$  is the degrees of freedom for the model
- ▶ GWEN:
  - $\hat{\mathbf{y}} = \mathbf{H}^*\mathbf{y} - \mathbf{T}$
  - Where  $T_i = \left\{ \mathbf{W}_i^{1/2} \mathbf{X} (\mathbf{X}' \mathbf{W}_i \mathbf{X} + (1 - \alpha) \lambda I_p)^{-1} \frac{\lambda_i}{\gamma_i} \right\}_i$
- ▶ GWEN is not a linear smoother
- ▶ Solution: use GWEN for selection then fit local model for the selected variables via traditional GWR

27

As in the case of traditional GWR, it is necessary to estimate the bandwidth parameter for a GWEN model.

For traditional GWR we can use the corrected AIC or generalized cross validation. Both methods use the trace of the projection matrix as the degrees of freedom for the model.

Because it incorporates an  $\mathcal{L}_1$  penalty, the fitted values from the GWEN are not a linear combination of the observed response. So the adaptive elastic net is not a linear smoother and so there is no projection matrix for a GWEN model.

For this reason, the GWEN is used for local variable selection and the coefficients of the resulting local model are fit using weighted least squares, as in traditional GWR.

# 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
  - Counter with locally linear coefficients
- ▶ Augment with covariate-by-location interactions:

$$Z_i = \begin{pmatrix} \tilde{X}_i & L_i \tilde{X}_i & M_i \tilde{X}_i \end{pmatrix}$$

Where

- ▶  $\tilde{X}_i$  is the matrix of covariates selected for the model at location  $s_i$
- ▶  $L_i = \text{diag}\{s_{i',x} - s_{i,x}\}$  for  $i' = 1, \dots, n$
- ▶  $M_i = \text{diag}\{s_{i',y} - s_{i,y}\}$  for  $i' = 1, \dots, n$

28

The traditional GWR fits models with locally constant coefficients, and can be thought of as a Nadaraya-Watson kernel smoother for the regression coefficients. Such a smoother is known to exhibit a "boundary effect", meaning that the smoother produces estimates that are biased near the boundary of the domain, especially when there is a gradient at the boundary.

To reduce the boundary effect, the GWEN can be fit with coefficient estimates that are locally linear, rather than locally constant. This is done after variable selection by augmenting the selected covariates with covariate-by-location interactions.

The interactions are on both the x and y coordinates, so each covariate that is selected for the model will appear three times in the model.

Now I'd like to discuss a simulation study.

## Simulation study

29

New Section

# Simulation study

## Simulating covariates

- ▶  $30 \times 30$  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:  $\rho$ 
  - none ( $\rho = 0$ )
  - moderate ( $\rho = 0.5$ )

30

In order to assess the utility of the GWEN for variable selection and coefficient estimation in a varying coefficients model, I performed a simulation study.

Spatial data were simulated on a 30 by 30 grid covering the domain  $[0,1] \times [0,1]$ . Five covariates were simulated using Gaussian random fields with an exponential covariance function.

The marginal variance of the covariates was one, and the range of the covariance function was 0.1.

The covariates were simulated at two levels of collinearity: none, and moderate, for which the correlation was set to 0.5

# Simulation study

## Simulating the response

- ▶  $Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\boldsymbol{\beta}(\mathbf{s}) = \sum_{j=1}^5 \beta_j(\mathbf{s})X_j(\mathbf{s}) + \varepsilon(\mathbf{s})$
- ▶  $\beta_1(\mathbf{s})$ , the coefficient function for  $X_1$ , is nonzero in part of the domain.
- ▶ Coefficients for  $X_2, \dots, X_5$  are zero everywhere
- ▶  $\varepsilon(\mathbf{s}) \sim iid N(0, \sigma^2)$ 
  - Low noise:  $\sigma = 0.5$
  - High noise:  $\sigma = 1$

31

With the covariates in hand, the response variable was generated via a linear model.

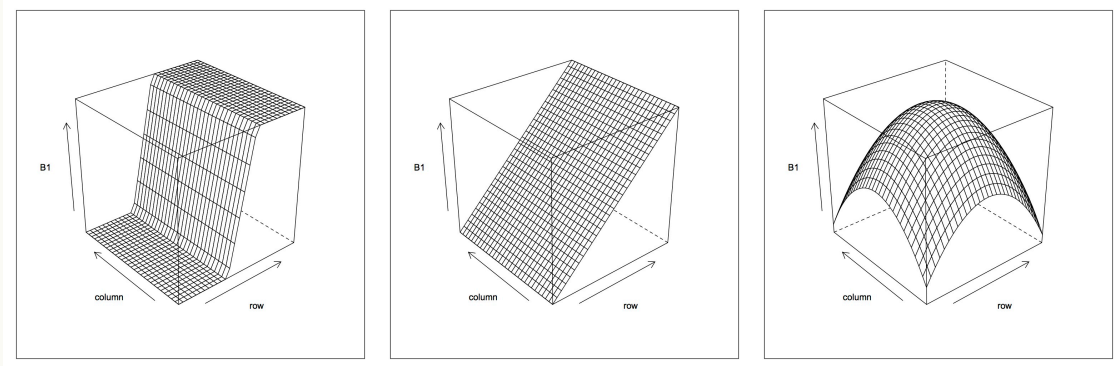
There are five covariates; four of them ( $\beta_2$  through  $\beta_5$ ) have a coefficient of zero everywhere on the domain.

For all the simulation settings, the coefficient of  $\beta_1$  ranges within the domain from a minimum of zero to a maximum of one.

The random noise added to the linear model was iid Gaussian noise at two different settings for the variance. The low-noise setting was  $\sigma = 0.5$  and the high-noise setting was  $\sigma = 1$ .

# Simulation study

Coefficient functions: step, gradient, and parabola



32

The response variable was simulated for three different types of coefficient surface  $\beta_1$ .

First is a step function where the "step" is on a slope rather than a discontinuity.

Second is a constant gradient.

Third is a parabola centered at the center of the domain.



## Simulation study

Simulation settings

Each setting simulated 100 times:

Setting	function	$\rho$	$\sigma^2$
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

33

The table lists the twelve settings for the simulation study, where once again the parameters being varied across the settings are the coefficient surface  $\beta_1$ , the amount of colinearity in the covariates, and the variance of the random noise.

Each setting was simulated 100 times and each time, and...

## Simulation study

### Estimation methods

method	selection	locally linear
GWR	NA	
oracular GWR	oracle	x
GWEN	GWEN	
GWAL	GWAL	
GWEN-LLE	GWEN	x
GWAL-LLE	GWAL	x

34

The following methods were used to estimate the varying coefficient regression model:

traditional GWR

oracular GWR (with locally linear estimation of the coefficients)

the GWEN with locally constant coefficients

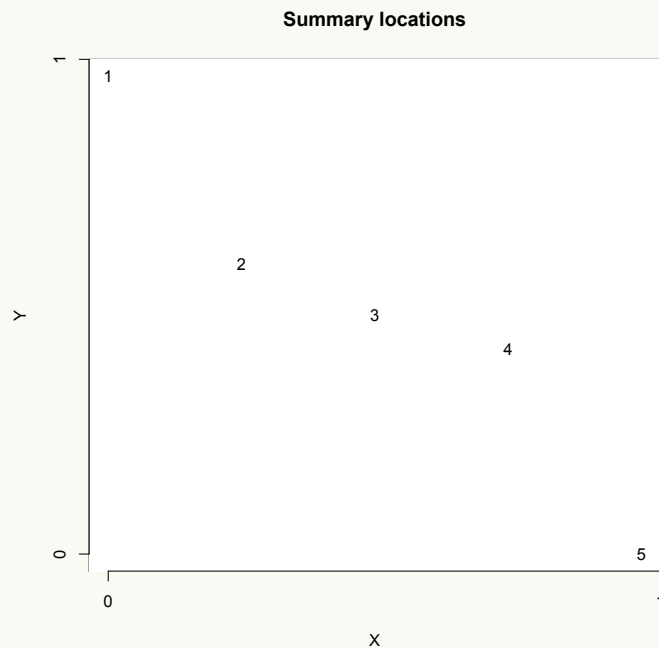
the GWAL with locally constant coefficients

the GWEN with locally linear coefficients

the GWAL with locally linear coefficients.

# Simulation results

## Summary locations



35

The results of the simulation were summarized at these five locations.

Locations two and four are at the "corners" of the step function.

The correct result of selection is ambiguous at some locations where the summary location is at the spot where the true coefficient surface  $\beta_4$  changes from zero to nonzero.

In particular, selection is ambiguous at location four for the step function, at location five for the gradient, and at locations one and five for the parabola.

We'll ignore selection accuracy of  $\beta_1$  where that is ambiguous

First we will consider the variable selection results of the simulation.

# Simulation results

## Selection performance

- ▶ GWEN selection (60 cases):
  - 21 with no false positives
  - 30 with no false negatives
  - 13 with neither
- ▶ GWAL selection (60 cases):
  - 27 with no false positives
  - 26 with no false negatives
  - 17 with neither
- ▶ Selection errors almost always below five percent
- ▶ Worst false positive rate: 8% at location three of the step function (GWEN selection)
- ▶ Worst false negative rate: 13% (same location, GWAL selection)
- ▶ No consistent difference between GWEN and GWAL

36

The GWEN and the GWAL seem to perform well in selection in the simulation.

considering each simulation case to be a specific location for a specific setting,

the GWEN produced 13 cases out of 60 with no selection errors over 100 simulations. The GWAL had 17 cases with no selection errors.

Increased noise variance and Increased colinearity both led to worse selection performance

Several of the cases without selection errors occurred for the settings with moderate colinearity. Only four of the cases without selection errors occurred in settings with high noise variance.

# Simulation results

## Estimation performance

- ▶ Cases where oracular GWR clearly had the best performance
  - Minimum  $\text{MSE}(\hat{\beta}_1)$ : 38 of 60
  - Minimum  $\text{var}(\hat{\beta}_1)$ : 44 of 60
- ▶ Generally small differences between GWR, oracular GWR, GWEN-LLE, and GWAL-LLE
- ▶ Methods with locally constant coefficients had larger bias at the boundaries
- ▶ Fitting  $\hat{y}$ : MSE nearest  $\sigma^2$  split between GWAL-LLE, oracle, and GWR

37

The method of oracular selection produced the best performance in terms of the MSE in estimating  $\beta_1$ , producing the smallest MSE in 38 of the 60 summarized simulation locations. It also had the lowest variance of estimating  $\beta_1$  for 44 of 60 cases.

In most cases the estimation accuracy in terms of MSE was not much different between GWR, oracular GWR, GWEN-LLE, and GWAL-LLE.

At the boundaries of the domain, especially for the parabola coefficient function, the lower bias of the locally linear estimators led them to have lower MSE than the methods that fit locally constant coefficients.

Increased noise variance led to worse estimation accuracy

Increased collinearity in the covariates led to worse estimation accuracy

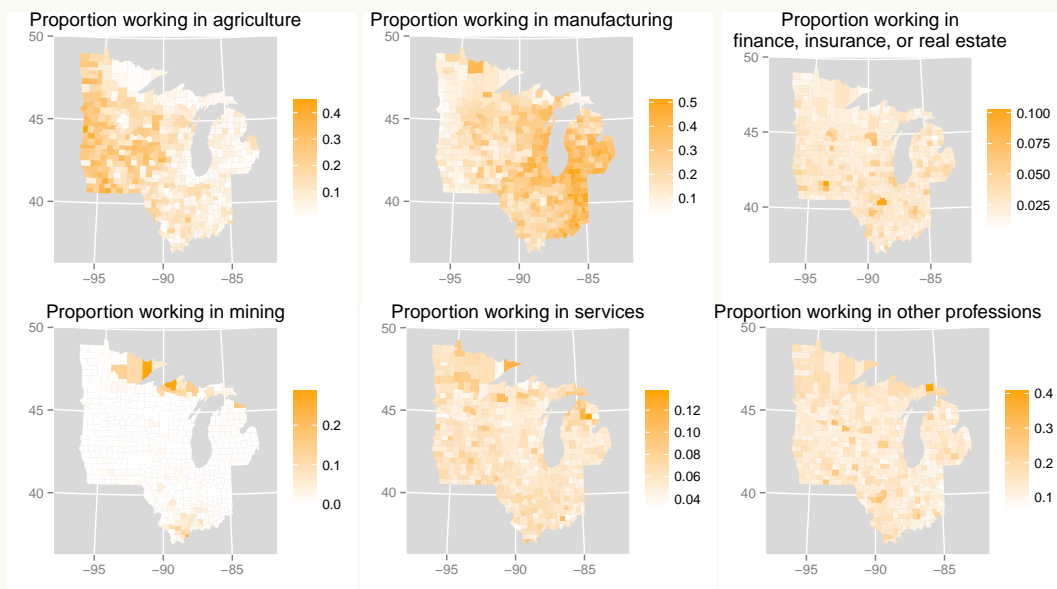
Data example: poverty rate in the upper  
midwest

38

New Section

# Data example: poverty rate in the upper midwest

## Revisiting the motivating example



39

We return to the data example that was introduced at the beginning of the talk.

Again, these are the covariates for a model of the county-level poverty rate.

## 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

40

The covariates are the proportion of the county's population working in the economic sectors of agriculture, finance, manufacturing, mining, services, and other professions.

The response variable in this model is the logit-transformed county-level poverty rate from the 1970 US census.

The model's domain is the upper midwest states of Minnesota, Iowa, Wisconsin, Illinois, Indiana, and Michigan.



# 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

41

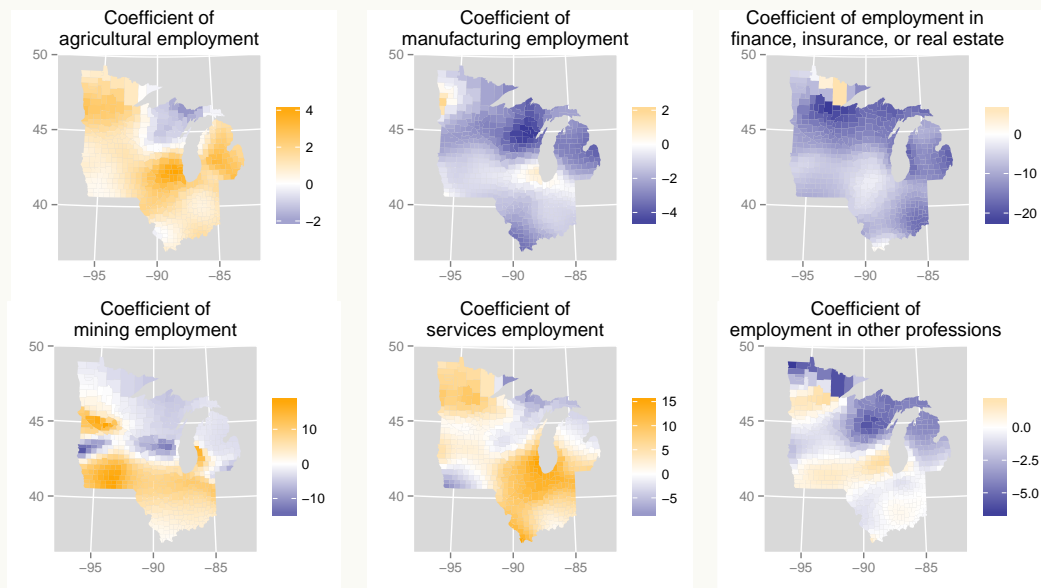
Since data is aggregated on counties and the counties exactly divide the domain, this is actually areal data.

We'll treat it as geostatistical data by assuming each county's data is sampled at the county's centroid.

The data is modeled using both traditional GWR and the GWEN with locally linear coefficient estimates.

# Data example: poverty rate in the upper midwest

## Results from traditional GWR

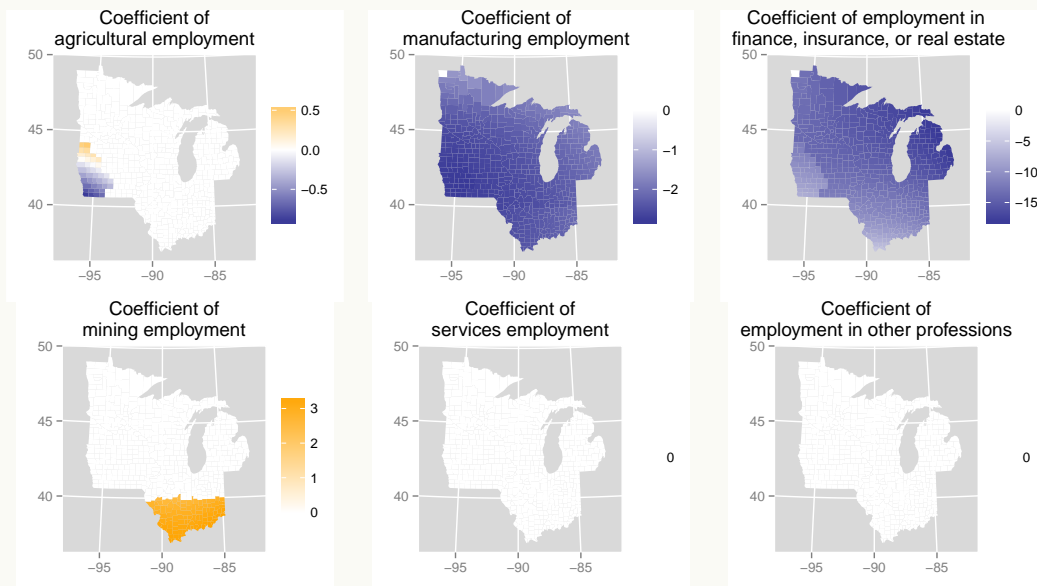


42

Here are plots of the coefficient estimates from the model fit by traditional GWR.

# Data example: poverty rate in the upper midwest

## Results from GWEN



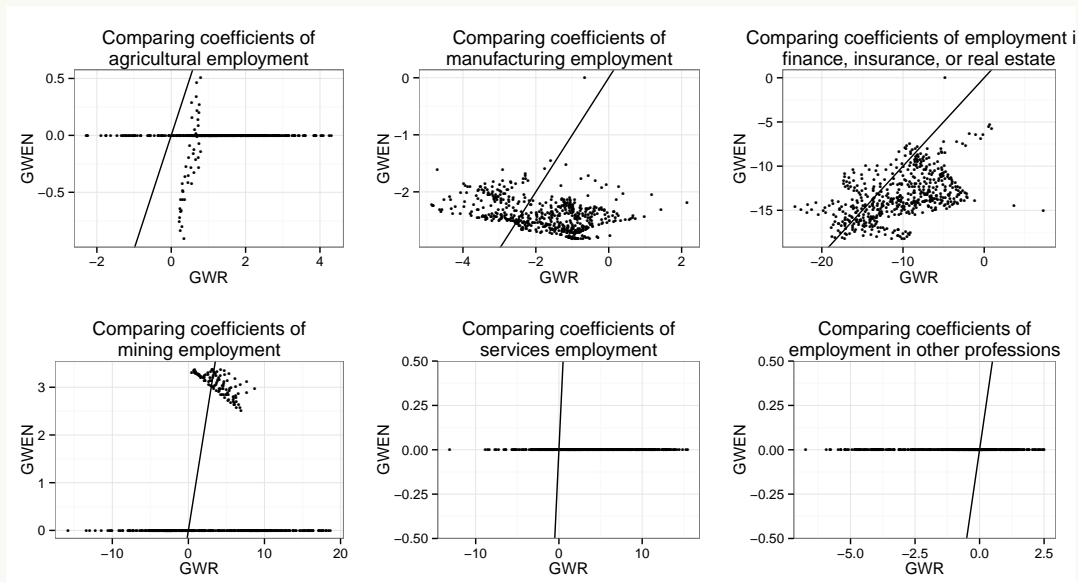
43

These are plots of the coefficients estimates from the model fit by the GWEN with locally linear coefficients.

The GWEN has selected just a few covariates as being important predictors of the county level poverty rate.

# Data example: poverty rate in the upper midwest

## Comparing the coefficients from GWR and the GWEN



44

This slide plots the estimates from the GWEN against those from traditional GWR. Each plot has a 1-1 line, along which the data points would lie if the GWEN and GWR produced identical results.

The GWEN resulted in coefficient estimates that are less variable than traditional GWR.

## Data example: poverty rate in the upper midwest

Results from GWEN-LLE

- ▶ Relatively constant compared to GWR
- ▶ Not associated with poverty rate: Services, "other professions" sector
- ▶ 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

Some observations about the models produced by traditional GWR and the GWEN-LLE:

per the GWEN, Employment in services and the "other professions" sectors did not affect the poverty rate anywhere.

Employment in manufacturing and finance were both associated with a decreased poverty rate across the entire domain. This was not dissimilar to the relationship estimated by traditional GWR.

Agricultural employment was selected as a meaningful predictor of the poverty rate only in western Iowa. Within that region there was a north-south gradient to the coefficient.

## Future work

46

New Section

## Future work

- ▶ Apply the GWEN to models for non-Gaussian response variable
- ▶ Investigate the asymptotic properties of the GWEN
- ▶ Incorporate spatial autocorrelation in the model
- ▶ PalEON project: modeling and mapping tree biomass in the upper midwest

note

Thank you!

Thank you all, that concludes my presentation. I'd like to especially acknowledge Jun for her patient work advising my research, and Katherine for her collaboration on the analysis of the county-level poverty data.