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

Wesley Brooks

Department of Statistics
University of Wisconsin–Madison

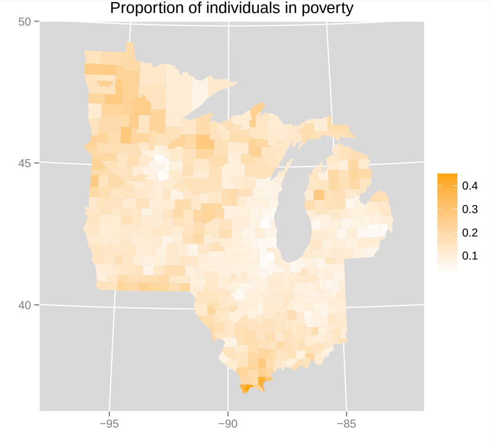
somesquares.org

# Section 1

# Motivation

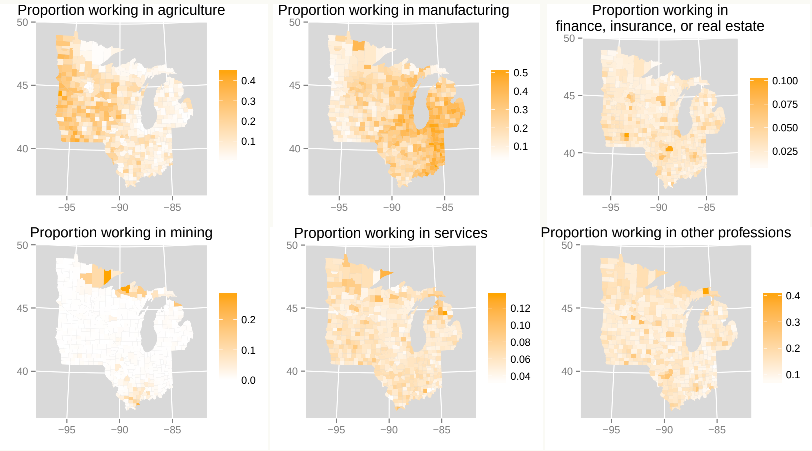
# Motivation

Take a look at some data



# Motivation

Take a look at some data



# Motivation

Sensible questions about the data

- Which of the economic-structure variables is associated with poverty rate?

# Motivation

Sensible questions about the data

- Which of the economic-structure variables is associated with poverty rate?
- What are the sign and magnitude of that association?

# Motivation

Sensible questions about the data

- 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?

# Motivation

Sensible questions about the data

- 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?
- Are the sign and magnitude of the associations constant across the region?

# Section 2

# Introduction

# Introduction

A review of existing methods

- Spatial regression

# Introduction

A review of existing methods

- Spatial regression
- Varying coefficient regression

# Introduction

A review of existing methods

- Spatial regression
- Varying coefficient regression
  - Splines

# Introduction

A review of existing methods

- Spatial regression
- Varying coefficient regression
  - Splines
  - Kernels

# Introduction

A review of existing methods

- Spatial regression
- Varying coefficient regression
  - Splines
  - Kernels
  - Wavelets

# Introduction

A review of existing methods

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

# Introduction

Some definitions

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

# Introduction

Further definitions

- Geostatistical data:
  - Observations are made at sampling locations $\boldsymbol{s}_i$ for $i = 1, \ldots, n$
  - E.g. elevation, temperature
- Areal data:
  - Domain is partitioned into $n$ regions $\{D_1, \ldots, D_n\}$
  - The regions do not overlap, and they divide the domain completely: $\mathcal{D} = \bigcup_{i=1}^{n} D_i$
  - Sampling locations $\boldsymbol{s}_i$ for $i = 1, \ldots, n$ are the centroids of the regions
  - E.g. poverty rate, population, spatial mean temperature

# Introduction

Existing approaches: spatial regression

- The typical spatial regression (Cressie, 1993)

$$\begin{aligned} Y(\boldsymbol{s}) &= \boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta} + W(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \\ \text{cov}(W(\boldsymbol{s}), W(\boldsymbol{t})) &= \Gamma\left(\delta(\boldsymbol{s}, \boldsymbol{t})\right) \\ \delta(\boldsymbol{s}, \boldsymbol{t}) &= \sqrt{\|\boldsymbol{t} - \boldsymbol{s}\|_2} \\ \text{E.g. } \Gamma(\delta(\boldsymbol{s}, \boldsymbol{t})) &= \exp\{-\phi^{-1}\delta(\boldsymbol{s}, \boldsymbol{t})\} \qquad (1) \end{aligned}$$

# Introduction

Existing approaches: spatial regression

- The typical spatial regression (Cressie, 1993)

$$
\begin{aligned}
Y(\boldsymbol{s}) &= \boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta} + W(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \\
\text{cov}(W(\boldsymbol{s}), W(\boldsymbol{t})) &= \Gamma\left(\delta(\boldsymbol{s}, \boldsymbol{t})\right) \\
\delta(\boldsymbol{s}, \boldsymbol{t}) &= \sqrt{\|\boldsymbol{t} - \boldsymbol{s}\|_2} \\
\text{E.g. } \Gamma(\delta(\boldsymbol{s}, \boldsymbol{t})) &= \exp\{-\phi^{-1}\delta(\boldsymbol{s}, \boldsymbol{t})\}
\end{aligned}
\tag{1}
$$

- $W(\boldsymbol{s})$ is a spatial random effect that accounts for autocorrelation in the response variable

# Introduction

Existing approaches: spatial regression

- The typical spatial regression (Cressie, 1993)

$$
\begin{aligned}
Y(\boldsymbol{s}) &= \boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta} + W(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \\
\text{cov}(W(\boldsymbol{s}), W(\boldsymbol{t})) &= \Gamma\left(\delta(\boldsymbol{s}, \boldsymbol{t})\right) \\
\delta(\boldsymbol{s}, \boldsymbol{t}) &= \sqrt{\|\boldsymbol{t} - \boldsymbol{s}\|_2} \\
\text{E.g. } \Gamma(\delta(\boldsymbol{s}, \boldsymbol{t})) &= \text{exp}\{-\phi^{-1}\delta(\boldsymbol{s}, \boldsymbol{t})\}
\end{aligned}
\tag{1}
$$

- $W(\boldsymbol{s})$ is a spatial random effect that accounts for autocorrelation in the response variable
- The coefficients $\boldsymbol{\beta}$ are constant

# Introduction

Existing approaches: spatial regression

- The typical spatial regression (Cressie, 1993)

$$
\begin{aligned}
Y(\boldsymbol{s}) &= \boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta} + W(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \\
\text{cov}(W(\boldsymbol{s}), W(\boldsymbol{t})) &= \Gamma\left(\delta(\boldsymbol{s}, \boldsymbol{t})\right) \\
\delta(\boldsymbol{s}, \boldsymbol{t}) &= \sqrt{\|\boldsymbol{t} - \boldsymbol{s}\|_2} \\
\text{E.g. } \Gamma(\delta(\boldsymbol{s}, \boldsymbol{t})) &= \text{exp}\{-\phi^{-1}\delta(\boldsymbol{s}, \boldsymbol{t})\}
\end{aligned} \tag{1}
$$

- $W(\boldsymbol{s})$ is a spatial random effect that accounts for autocorrelation in the response variable
- The coefficients $\boldsymbol{\beta}$ are constant
- Relies on *a priori* global variable selection

# Introduction

Existing approaches: varying coefficients regression

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

# Introduction

Existing approaches: varying coefficients regression

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

- Assume an effect modifying variable $S$

# Introduction

Existing approaches: varying coefficients regression

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

- Assume an effect modifying variable $S$
- Coefficients are functions of $S$

# Introduction

Existing approaches: varying coefficients regression

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

- Assume an effect modifying variable $S$
- Coefficients are functions of $S$
- Generally assume that the coefficient functions are smooth

# Introduction

Existing approaches: varying coefficients regression

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

- Assume an effect modifying variable $S$
- Coefficients are functions of $S$
- Generally assume that the coefficient functions are smooth
- This is a varying coefficient regression (VCR) model (Hastie and Tibshirani, 1993)

# Introduction

Existing approaches: varying coefficients regression

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

- Assume an effect modifying variable $S$
- Coefficients are functions of $S$
- Generally assume that the coefficient functions are smooth
- This is a varying coefficient regression (VCR) model (Hastie and Tibshirani, 1993)
- If $s$ is a spatial location then we have a spatially varying coefficient regression (SVCR) model

# Introduction

Existing approaches: spatially varying coefficient process

- Making model more flexible: coefficients in a spatial regression model can be allowed to vary (Gelfand et al., 2003)

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

# Introduction

Existing approaches: spatially varying coefficient process

- Making model more flexible: coefficients in a spatial regression model can be allowed to vary (Gelfand et al., 2003)

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

- The spatial random effect has been incorporated into the spatially varying intercept

# Introduction

Existing approaches: spatially varying coefficient process

- Making model more flexible: coefficients in a spatial regression model can be allowed to vary (Gelfand et al., 2003)

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

- The spatial random effect has been incorporated into the spatially varying intercept
- $\{\beta_1(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}, \ldots, \{\beta_p(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ are stationary spatial processes with Matèrn covariance functions

# Introduction

Existing approaches: spatially varying coefficient process

- Making model more flexible: coefficients in a spatial regression model can be allowed to vary (Gelfand et al., 2003)

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

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

# Introduction

Existing approaches: spline-based VCR and SVCR models

- Splines are a way to parameterize smooth functions
- Splines can be incorporated into a generalized additive model (GAM):
  - $E\{Y(t)\} = f\{X_1(t)\} + \cdots + f\{X_p(t)\}$
- It is possible to parameterize a VCR model with splines for the coefficient functions:
  - $E\{Y(t)\} = \beta_1(t)X_1(t) + \cdots + \beta_p(t)X_p(t)$

# Introduction

Existing approaches: 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 (maximize the likelihood plus a penalty) to encourage coefficient functions to be zero
- SCAD penalty (Fan and R. Li, 2001) on the integral of the square of the coefficient function (Wang, H. Li, and Huang, 2008)
- Non-negative garrote penalty (Breiman, 1995) on the integral of the square of the coefficient function (Antoniadas, Gijbels, and Verhasselt, 2012)

# Introduction

Existing approaches: wavelet methods for VCR models

Wavelet methods involve decomposing a function into local frequency components. Wavelet methods for VCR models include using Bayesian variable selection or the Lasso to estimate which local frequency components have nonzero coefficients (Shang, 2011; J. Zhang and Clayton, 2011).

These methods achieve sparsity in the local frequency components but not in the local covariates, and so are not suitable for local variable selection.

# Section 3

# Geographically weighted regression

# Geographically weighted regression

Existing approaches: geographically weighted regression

When the effect modifying variable $s$ refers to spatial location, the method of local regression is called geographically weighted regression (GWR) (Brundson, S. Fotheringham, and Martin Charlton, 1998; A. Fotheringham, Brunsdon, and M. Charlton, 2002)

# Geographically weighted regression

Existing approaches: Local regression

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

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

Where $\phi$ is a bandwidth parameter.
Given the weights, a local model is fit at each sampling location using the local likelihood (Loader, 1999)

# Geographically weighted regression

Existing approaches: Local likelihood

Calibrate the model by doing the following at each sampling location:

- Weight each observation's likelihood
- Weights are given by the kernel

$$
\begin{aligned}
L &= \prod_{i'=1}^{n} \left(L_{i'}\right)^{w_{ii'}} \\
\ell &= \sum_{i'=1}^{n} w_{ii'} \left\{ \text{log}\, \sigma_i^2 + \sigma_i^{-2} \left(y_{i'} - \boldsymbol{x}_{i'}' \boldsymbol{\beta}_i\right)^2 \right\}
\end{aligned}
$$

Where $\boldsymbol{\beta}_i = \boldsymbol{\beta}(\boldsymbol{s}_i)$.

# Geographically weighted regression

Existing approaches: bandwidth estimation for GWR

- Smaller bandwidth: less bias, more flexible coefficient surface
- Large bandwidth: less variance, less flexible coefficient surface
- Estimate the degrees of freedom used in estmating the coefficient surface (Hurvich, Simonoff, and Tsai, 1998):
  - $\hat{\boldsymbol{y}} = H\boldsymbol{y}$
  - $\nu = \mathrm{tr}(H)$
- Then the corrected AIC for bandwidth selection is:
- $\mathrm{AIC_c} = 2n \log \sigma + n \left\{ \frac{n+\nu}{n-2-\nu} \right\}$

# Geographically weighted regression

Existing approaches: geographically weighted Lasso

Within a GWR model, using the Lasso (Tibshirani, 1996) for local variable selection is called the geographically weighted Lasso (GWL) (Wheeler, 2009).

- 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 jacknife criterion: $|y_i - \hat{y}_i|$
- The jacknife 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

# Section 4

# Local variable selection and parameter estimation

# Local variable selection and parameter estimation

Geographically weighted 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

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

# Local variable selection and parameter estimation

Geographically weighted elastic net (GWEN)

where $\sum_{i'=1}^{n} w_{ii'} \left(y_{i'} - \boldsymbol{x}'_{i'}\boldsymbol{\beta}_i\right)^2$ is the weighted sum of squares minimized by traditional GWR, and $\mathcal{J}_2(\boldsymbol{\beta}_i) = \alpha_i \lambda_i^* \sum_{j=1}^{p} |\beta_{ij}|/\gamma_{ij} + (1-\alpha_i)\lambda_i^* \sum_{j=1}^{p} \left(\beta_{ij}/\gamma_{ij}\right)^2$ is the AEN penalty.

# Local variable selection and parameter estimation

Geographically weighted elastic net (GWEN)

It is necessary to estimate an AEN tuning parameter for each local model. Using the local BIC allows fitting a local model at any location within the domain

$$
\begin{aligned}
\mathrm{BIC}_{\mathrm{loc},i} &= -2\sum_{i'=1}^{n} \ell_{ii'} + \log\left(\sum_{i'=1}^{n} w_{ii'}\right) \mathrm{df}_i \\
&= -2\sum_{i'=1}^{n} \log\left[\left(2\pi\hat{\sigma}_i^2\right)^{-1/2} \exp\left\{-\frac{1}{2}\hat{\sigma}_i^{-2}\left(y_{i'} - \boldsymbol{x}_{i'}'\hat{\boldsymbol{\beta}}_{i'}\right)^2\right\}\right]^{w} \\
&+ \log\left(\sum_{i'=1}^{n} w_{ii'}\right) \mathrm{df}_i
\end{aligned}
\qquad (3)
$$

# Local variable selection and parameter estimation

Geographically weighted elastic net (GWEN)

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

$$+ \log \left( \sum_{i'=1}^{n} w_{ii'} \right) \mathrm{df}_i$$

# Section 5

# Simulation study

# Simulation study

Simulating covariates

Five covariates $\tilde{X}_1, \ldots, \tilde{X}_5$ were simulated by Gaussian random fields on the domain $[0,1] \times [0,1]$ on a $30 \times 30$ grid of sampling locations:

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

Where the covariates were simulated with colinearity, the colinearity was induced by multiplying the design matrix by the square root of the colinearity matrix:

$$\begin{aligned} \text{diag}(\Omega_{5\times 5}) &= 1 \\ \text{off-diag}(\Omega_{5\times 5}) &= \rho \\ X &= \tilde{X}R \end{aligned} \tag{4}$$

Where $\Omega_{5\times 5} = R'R$ is the Cholesky decomposition.

# Simulation study

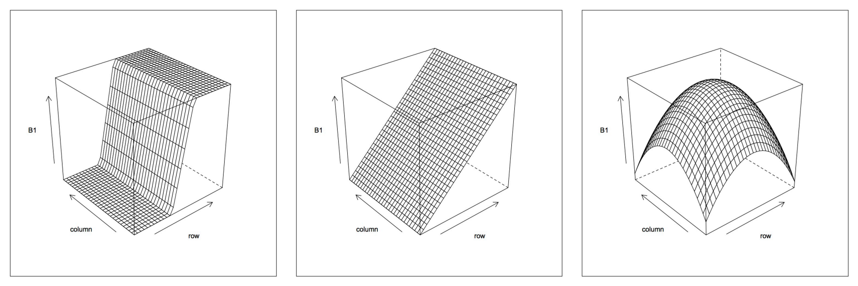
Simulating the response

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

# Simulation study

Coefficient functions

Call these functions step, gradient, and parabola:



# Simulation study

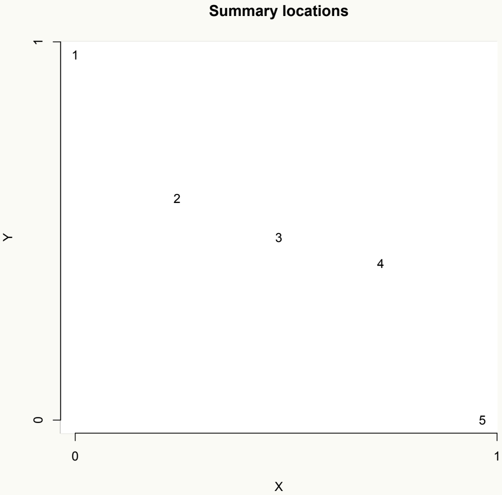
Simulation settings

| 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 |

Table : Simulation parameters for each setting.

# Simulation results

Selection



# Simulation results

## Selection

| | step | | | | gradient | | | | parabola | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GWEN | | GWAL | | GWEN | | GWAL | | GWEN | | GWAL | |
| location | $\beta_1$ | $\beta_2$ - $\beta_5$ | $\beta_1$ | $\beta_2$ - $\beta_5$ | $\beta_1$ | $\beta_2$ - $\beta_5$ | $\beta_1$ | $\beta_2$ - $\beta_5$ | $\beta_1$ | $\beta_2$ - $\beta_5$ | $\beta_1$ | $\beta_2$ - $\beta_5$ |
| 1 | 0.99 | 0.00 | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.36 | 0.00 | 0.38 | 0.00 |
| | 0.99 | 0.02 | 0.99 | 0.02 | 1.00 | 0.01 | 1.00 | 0.01 | 0.71 | 0.02 | 0.70 | 0.02 |
| | 0.99 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.28 | 0.00 | 0.33 | 0.00 |
| | 0.96 | 0.05 | 0.91 | 0.04 | 0.99 | 0.03 | 0.99 | 0.01 | 0.56 | 0.02 | 0.55 | 0.02 |
| 2 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 1.00 | 0.03 | 1.00 | 0.03 | 1.00 | 0.02 | 1.00 | 0.02 | 1.00 | 0.02 | 0.99 | 0.01 |
| | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.01 | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 0.99 | 0.05 | 0.97 | 0.04 | 1.00 | 0.02 | 0.99 | 0.01 | 0.98 | 0.02 | 0.97 | 0.01 |
| 3 | 0.91 | 0.01 | 0.91 | 0.00 | 1.00 | 0.01 | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 0.96 | 0.05 | 0.96 | 0.05 | 1.00 | 0.01 | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 0.92 | 0.05 | 0.95 | 0.02 | 1.00 | 0.02 | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 0.92 | 0.08 | 0.87 | 0.05 | 1.00 | 0.02 | 0.98 | 0.02 | 0.99 | 0.01 | 0.99 | 0.01 |
| 4 | 0.48 | 0.01 | 0.43 | 0.01 | 1.00 | 0.01 | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 0.72 | 0.04 | 0.78 | 0.03 | 1.00 | 0.01 | 1.00 | 0.01 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 0.49 | 0.02 | 0.46 | 0.02 | 1.00 | 0.02 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | 0.60 | 0.05 | 0.56 | 0.04 | 1.00 | 0.03 | 0.98 | 0.02 | 1.00 | 0.01 | 0.98 | 0.02 |
| 5 | 0.00 | 0.00 | 0.00 | 0.00 | 0.83 | 0.00 | 0.82 | 0.00 | 0.32 | 0.00 | 0.32 | 0.00 |
| | 0.03 | 0.01 | 0.02 | 0.00 | 0.70 | 0.00 | 0.66 | 0.00 | 0.68 | 0.02 | 0.73 | 0.02 |
| | 0.00 | 0.00 | 0.00 | 0.00 | 0.87 | 0.01 | 0.87 | 0.00 | 0.37 | 0.00 | 0.42 | 0.00 |
| | 0.06 | 0.02 | 0.01 | 0.02 | 0.61 | 0.01 | 0.62 | 0.02 | 0.61 | 0.04 | 0.58 | 0.03 |

Table : Selection frequency for the indicated variables.

# Simulation results

MSE of $\beta_1(s)$ - step coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| step | 1 | 0.026 | *0.025* | 0.057 | 0.057 | 0.062 | **0.008** |
| | | 0.042 | *0.040* | 0.193 | 0.180 | 0.102 | **0.016** |
| | | 0.036 | **0.014** | 0.055 | 0.067 | 0.080 | *0.016* |
| | | *0.093* | 0.130 | 0.230 | 0.285 | 0.144 | **0.030** |
| | 2 | 0.063 | 0.058 | 0.043 | *0.043* | **0.038** | 0.055 |
| | | 0.087 | 0.084 | **0.064** | *0.064* | 0.073 | 0.084 |
| | | 0.068 | 0.049 | 0.045 | *0.040* | **0.036** | 0.052 |
| | | 0.140 | 0.128 | *0.082* | 0.093 | **0.074** | 0.096 |
| | 3 | 0.025 | 0.025 | 0.019 | 0.019 | **0.004** | *0.010* |
| | | 0.021 | 0.021 | 0.015 | 0.015 | **0.007** | *0.011* |
| | | 0.027 | 0.021 | 0.018 | *0.014* | **0.006** | 0.019 |
| | | 0.027 | 0.038 | 0.020 | 0.031 | **0.007** | *0.016* |
| | 4 | *0.026* | 0.026 | 0.028 | **0.025** | 0.034 | 0.054 |
| | | **0.046** | *0.050* | 0.054 | 0.057 | 0.073 | 0.081 |
| | | **0.025** | 0.030 | 0.030 | *0.027* | 0.036 | 0.063 |
| | | **0.035** | *0.036* | 0.043 | 0.046 | 0.072 | 0.083 |
| | 5 | **0.000** | **0.000** | **0.000** | **0.000** | **0.000** | 0.008 |
| | | 0.002 | 0.002 | 0.001 | *0.000* | **0.000** | 0.014 |
| | | **0.000** | **0.000** | **0.000** | **0.000** | **0.000** | 0.021 |
| | | 0.006 | *0.004* | 0.016 | 0.009 | **0.000** | 0.029 |

Table : Mean squared error of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

MSE of $\beta_1(s)$ - gradient coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| | | 0.126 | 0.125 | **0.005** | *0.005* | 0.006 | 0.024 |
| | 1 | 0.105 | 0.102 | *0.026* | 0.027 | **0.019** | 0.042 |
| | | 0.136 | 0.132 | 0.005 | *0.005* | **0.005** | 0.029 |
| | | 0.135 | 0.119 | *0.043* | 0.044 | **0.023** | 0.055 |
| | | 0.006 | 0.006 | 0.001 | *0.001* | **0.001** | 0.002 |
| | 2 | 0.006 | 0.006 | 0.002 | *0.002* | **0.002** | 0.003 |
| | | 0.008 | 0.006 | *0.001* | 0.001 | **0.001** | 0.003 |
| | | 0.009 | 0.011 | *0.004* | 0.008 | **0.003** | 0.006 |
| | | 0.002 | 0.002 | **0.000** | 0.000 | *0.000* | 0.002 |
| | | 0.003 | 0.003 | 0.001 | *0.001* | **0.001** | 0.003 |
| gradient | 3 | 0.002 | 0.002 | *0.000* | 0.000 | **0.000** | 0.003 |
| | | 0.006 | 0.010 | *0.002* | 0.007 | **0.001** | 0.007 |
| | | 0.005 | 0.005 | *0.000* | **0.000** | 0.001 | 0.002 |
| | 4 | 0.007 | 0.006 | 0.002 | *0.002* | **0.002** | 0.004 |
| | | 0.004 | 0.005 | *0.000* | 0.000 | **0.000** | 0.002 |
| | | 0.009 | 0.010 | *0.003* | 0.006 | **0.002** | 0.009 |
| | | 0.108 | 0.110 | 0.002 | *0.002* | **0.000** | 0.022 |
| | 5 | 0.084 | 0.084 | 0.011 | *0.010* | **0.000** | 0.044 |
| | | 0.107 | 0.119 | *0.003* | 0.003 | **0.000** | 0.028 |
| | | 0.065 | 0.076 | *0.008* | 0.010 | **0.000** | 0.056 |

Table : Mean squared error of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

MSE of $\beta_1(s)$ - parabola coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| | | 0.050 | 0.054 | 0.019 | *0.017* | **0.001** | 0.123 |
| | 1 | 0.145 | 0.151 | 0.053 | *0.053* | **0.001** | 0.248 |
| | | 0.029 | 0.046 | 0.016 | *0.015* | **0.001** | 0.133 |
| | | 0.105 | 0.125 | *0.065* | 0.082 | **0.001** | 0.248 |
| | | **0.103** | 0.104 | 0.105 | 0.106 | *0.104* | 0.105 |
| | 2 | 0.088 | 0.091 | *0.085* | 0.086 | **0.079** | 0.086 |
| | | **0.092** | 0.100 | *0.099* | 0.100 | 0.100 | 0.104 |
| | | *0.085* | 0.097 | 0.091 | 0.103 | **0.077** | 0.094 |
| | | *0.148* | 0.150 | 0.156 | 0.157 | 0.156 | **0.144** |
| | | 0.110 | 0.114 | 0.121 | 0.126 | *0.108* | **0.086** |
| parabola | 3 | **0.139** | *0.143* | 0.150 | 0.150 | 0.152 | 0.144 |
| | | *0.111* | 0.122 | 0.130 | 0.139 | 0.117 | **0.101** |
| | | **0.110** | 0.112 | 0.115 | 0.116 | 0.115 | *0.111* |
| | 4 | 0.092 | 0.093 | 0.094 | 0.095 | *0.085* | **0.085** |
| | | **0.104** | 0.111 | 0.113 | 0.114 | 0.114 | *0.109* |
| | | **0.080** | 0.100 | 0.094 | 0.108 | *0.088* | 0.097 |
| | | 0.044 | 0.047 | *0.014* | 0.016 | **0.001** | 0.123 |
| | 5 | 0.155 | 0.153 | 0.102 | *0.101* | **0.001** | 0.250 |
| | | 0.040 | 0.060 | *0.012* | 0.018 | **0.001** | 0.136 |
| | | 0.111 | 0.126 | *0.055* | 0.061 | **0.001** | 0.234 |

Table : Mean squared error of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

Variance of $\beta_1(\boldsymbol{s})$ - step coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| step | 1 | 0.023 | *0.023* | 0.057 | 0.058 | 0.063 | **0.009** |
| | | *0.036* | 0.036 | 0.195 | 0.180 | 0.098 | **0.016** |
| | | 0.027 | **0.013** | 0.056 | 0.068 | 0.080 | *0.016* |
| | | *0.059* | 0.100 | 0.226 | 0.276 | 0.145 | **0.030** |
| | 2 | 0.014 | 0.013 | **0.006** | *0.006* | 0.006 | 0.008 |
| | | 0.017 | 0.017 | 0.011 | *0.011* | **0.008** | 0.013 |
| | | 0.012 | 0.010 | *0.005* | **0.004** | 0.006 | 0.010 |
| | | 0.021 | 0.033 | 0.021 | 0.037 | **0.008** | *0.014* |
| | 3 | 0.022 | 0.023 | 0.019 | 0.019 | **0.004** | *0.009* |
| | | 0.021 | 0.021 | 0.014 | 0.014 | **0.005** | *0.008* |
| | | 0.024 | 0.021 | 0.018 | *0.014* | **0.005** | 0.016 |
| | | 0.024 | 0.036 | 0.020 | 0.032 | **0.005** | *0.014* |
| | 4 | 0.022 | 0.023 | 0.023 | 0.022 | **0.006** | *0.007* |
| | | 0.025 | 0.024 | 0.025 | 0.022 | **0.006** | *0.008* |
| | | 0.021 | 0.025 | 0.024 | 0.023 | **0.005** | *0.013* |
| | | 0.026 | 0.027 | 0.029 | 0.032 | **0.009** | *0.015* |
| | 5 | **0.000** | **0.000** | **0.000** | **0.000** | **0.000** | 0.007 |
| | | 0.002 | 0.002 | 0.001 | *0.000* | **0.000** | 0.014 |
| | | **0.000** | **0.000** | **0.000** | **0.000** | **0.000** | 0.021 |
| | | 0.006 | *0.004* | 0.016 | 0.009 | **0.000** | 0.029 |

Table : Variance of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

Variance of $\beta_1(\boldsymbol{s})$ - gradient coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| | | 0.009 | 0.009 | *0.005* | 0.005 | 0.006 | **0.004** |
| | 1 | *0.012* | 0.013 | 0.026 | 0.027 | 0.019 | **0.012** |
| | | 0.008 | 0.008 | 0.005 | *0.005* | **0.005** | 0.007 |
| | | *0.020* | 0.023 | 0.043 | 0.044 | 0.023 | **0.016** |
| | | 0.003 | 0.002 | 0.001 | *0.001* | **0.001** | 0.002 |
| | 2 | 0.004 | 0.004 | 0.002 | *0.002* | **0.002** | 0.003 |
| | | 0.003 | 0.003 | *0.001* | 0.001 | **0.001** | 0.003 |
| | | 0.005 | 0.008 | *0.004* | 0.008 | **0.003** | 0.006 |
| | | 0.002 | 0.002 | *0.000* | 0.000 | **0.000** | 0.002 |
| | | 0.003 | 0.003 | 0.001 | *0.001* | **0.001** | 0.003 |
| gradient | 3 | 0.002 | 0.002 | *0.000* | 0.000 | **0.000** | 0.003 |
| | | 0.006 | 0.010 | *0.002* | 0.007 | **0.001** | 0.007 |
| | | 0.003 | 0.003 | *0.000* | **0.000** | 0.000 | 0.002 |
| | 4 | 0.006 | 0.006 | 0.002 | *0.002* | **0.002** | 0.004 |
| | | 0.003 | 0.002 | *0.000* | 0.000 | **0.000** | 0.002 |
| | | 0.009 | 0.010 | *0.003* | 0.006 | **0.002** | 0.008 |
| | | 0.022 | 0.023 | 0.003 | *0.002* | **0.000** | 0.006 |
| | 5 | 0.029 | 0.032 | 0.011 | *0.010* | **0.000** | 0.011 |
| | | 0.018 | 0.019 | *0.003* | 0.003 | **0.000** | 0.009 |
| | | 0.030 | 0.033 | *0.008* | 0.011 | **0.000** | 0.017 |

Table : Variance of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

## Variance of $\beta_1(\boldsymbol{s})$ - parabola coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| | | 0.039 | 0.040 | 0.019 | 0.017 | **0.000** | *0.010* |
| | 1 | 0.057 | 0.059 | 0.047 | 0.047 | **0.000** | *0.019* |
| | | 0.026 | 0.037 | 0.016 | *0.015* | **0.000** | 0.018 |
| | | 0.060 | 0.071 | 0.063 | 0.074 | **0.000** | *0.031* |
| | | 0.003 | 0.003 | 0.003 | **0.003** | *0.003* | 0.003 |
| | 2 | *0.005* | 0.011 | 0.006 | 0.011 | 0.005 | **0.004** |
| | | 0.003 | 0.002 | *0.002* | 0.002 | **0.002** | 0.004 |
| | | 0.024 | 0.027 | 0.021 | 0.029 | **0.006** | *0.010* |
| | | *0.002* | 0.002 | 0.002 | 0.002 | 0.003 | **0.002** |
| | | 0.010 | 0.013 | 0.011 | 0.014 | **0.007** | *0.008* |
| parabola | 3 | *0.002* | **0.002** | 0.003 | 0.003 | 0.003 | 0.003 |
| | | 0.023 | 0.026 | 0.024 | 0.026 | **0.011** | *0.012* |
| | | 0.002 | 0.002 | **0.002** | *0.002* | 0.002 | 0.002 |
| | 4 | 0.006 | 0.006 | 0.007 | 0.008 | **0.005** | *0.005* |
| | | 0.003 | 0.002 | **0.002** | *0.002* | 0.002 | 0.003 |
| | | 0.012 | 0.021 | 0.010 | 0.024 | **0.007** | *0.009* |
| | | 0.036 | 0.038 | *0.014* | 0.016 | **0.000** | 0.015 |
| | 5 | 0.072 | 0.063 | 0.084 | 0.086 | **0.000** | *0.021* |
| | | 0.031 | 0.043 | *0.012* | 0.017 | **0.000** | 0.020 |
| | | 0.060 | 0.069 | 0.052 | 0.058 | **0.000** | *0.032* |

Table : Variance of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

Bias of $\beta_1(\boldsymbol{s})$ - step coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| | | -0.056 | -0.049 | **0.001** | *0.005* | 0.015 | -0.007 |
| | 1 | -0.080 | -0.069 | *0.020* | 0.040 | 0.072 | **0.002** |
| | | -0.093 | -0.037 | -0.010 | -0.009 | *-0.005* | **0.003** |
| | | -0.185 | -0.177 | -0.075 | -0.110 | *0.032* | **-0.009** |
| | | -0.222 | -0.213 | -0.193 | *-0.191* | **-0.178** | -0.217 |
| | 2 | -0.264 | -0.259 | **-0.231** | *-0.232* | -0.256 | -0.268 |
| | | -0.236 | -0.197 | -0.199 | *-0.188* | **-0.176** | -0.204 |
| | | -0.345 | -0.309 | *-0.248* | **-0.236** | -0.257 | -0.286 |
| | | -0.057 | -0.047 | *-0.006* | **-0.006** | 0.025 | 0.024 |
| step | 3 | *-0.009* | **0.004** | 0.022 | 0.024 | 0.047 | 0.051 |
| | | -0.052 | *-0.007* | **0.003** | 0.020 | 0.039 | 0.055 |
| | | -0.062 | -0.046 | **-0.011** | *-0.014* | 0.046 | 0.046 |
| | | 0.066 | *0.058* | 0.071 | **0.057** | 0.168 | 0.218 |
| | 4 | **0.147** | *0.165* | 0.170 | 0.188 | 0.260 | 0.272 |
| | | **0.062** | 0.071 | 0.077 | *0.067* | 0.174 | 0.223 |
| | | *0.098* | **0.098** | 0.121 | 0.119 | 0.250 | 0.262 |
| | | **0.000** | **0.000** | **0.000** | **0.000** | **0.000** | -0.022 |
| | 5 | 0.003 | *0.001* | -0.005 | -0.003 | **0.000** | -0.018 |
| | | **0.000** | **0.000** | **0.000** | **0.000** | **0.000** | 0.003 |
| | | 0.016 | *0.006* | -0.007 | 0.010 | **0.000** | 0.012 |

Table : Bias of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

Bias of $\beta_1(\boldsymbol{s})$ - gradient coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| | | -0.342 | -0.341 | *0.007* | 0.007 | **0.002** | -0.141 |
| | 1 | -0.305 | -0.299 | **0.019** | *0.021* | 0.024 | -0.174 |
| | | -0.357 | -0.352 | -0.009 | *-0.008* | **-0.003** | -0.151 |
| | | -0.340 | -0.311 | -0.026 | *-0.019* | **-0.007** | -0.197 |
| | | -0.061 | -0.061 | 0.001 | **0.000** | *-0.001* | -0.003 |
| | 2 | -0.049 | -0.046 | 0.004 | **0.004** | *0.004* | -0.011 |
| | | -0.073 | -0.062 | *-0.002* | **-0.001** | -0.002 | -0.003 |
| | | -0.064 | -0.054 | *0.004* | 0.004 | **0.003** | -0.008 |
| | | 0.003 | 0.006 | *-0.001* | -0.002 | -0.002 | **0.001** |
| gradient | 3 | **-0.000** | 0.003 | -0.005 | -0.003 | -0.003 | *0.001* |
| | | **-0.001** | 0.009 | 0.002 | 0.002 | *0.001* | 0.005 |
| | | -0.023 | -0.013 | **-0.002** | -0.009 | *0.003* | 0.011 |
| | | 0.047 | 0.050 | -0.003 | *-0.003* | -0.004 | **0.002** |
| | 4 | 0.028 | 0.032 | -0.010 | *-0.008* | **-0.007** | 0.011 |
| | | 0.043 | 0.055 | *0.003* | **0.003** | 0.003 | 0.010 |
| | | 0.008 | 0.025 | **-0.002** | -0.003 | *-0.002* | 0.029 |
| | | 0.293 | 0.296 | *-0.000* | 0.000 | **0.000** | 0.126 |
| | 5 | 0.235 | 0.228 | *0.013* | 0.017 | **0.000** | 0.182 |
| | | 0.298 | 0.318 | 0.003 | *0.002* | **0.000** | 0.137 |
| | | 0.189 | 0.208 | -0.004 | *-0.001* | **0.000** | 0.197 |

Table : Bias of $\hat{\beta}_1$ (**minimum**, *next best*).

# Simulation results

Bias of $\beta_1(s)$ - parabola coefficient surface

| function | location | GWEN | GWAL | GWEN-LLE | GWAL-LLE | oracle | GWR |
|---|---|---|---|---|---|---|---|
| | | 0.108 | 0.118 | **0.014** | *0.020* | -0.034 | 0.336 |
| | 1 | 0.299 | 0.303 | 0.082 | *0.081* | **-0.034** | 0.479 |
| | | 0.059 | 0.097 | **0.002** | *0.017* | -0.034 | 0.339 |
| | | 0.214 | 0.233 | *0.052* | 0.090 | **-0.034** | 0.466 |
| | | **0.316** | *0.318* | 0.319 | 0.321 | 0.318 | 0.319 |
| | 2 | 0.288 | 0.283 | 0.281 | *0.273* | **0.271** | 0.286 |
| | | **0.299** | 0.313 | *0.311* | 0.313 | 0.313 | 0.317 |
| | | **0.248** | 0.266 | *0.264* | 0.273 | 0.267 | 0.290 |
| | | *0.382* | 0.385 | 0.391 | 0.393 | 0.391 | **0.378** |
| | | *0.316* | 0.319 | 0.331 | 0.335 | 0.317 | **0.281** |
| parabola | 3 | **0.369** | 0.376 | 0.384 | 0.384 | 0.386 | *0.375* |
| | | **0.298** | 0.310 | 0.326 | 0.336 | 0.326 | *0.299* |
| | | **0.329** | 0.331 | 0.336 | 0.337 | 0.336 | *0.330* |
| | 4 | 0.294 | 0.294 | 0.295 | 0.295 | *0.284* | **0.282** |
| | | **0.318** | 0.329 | 0.333 | 0.335 | 0.335 | *0.327* |
| | | **0.262** | *0.281* | 0.290 | 0.290 | 0.285 | 0.297 |
| | | 0.090 | 0.094 | **0.001** | *0.006* | -0.034 | 0.329 |
| | 5 | 0.289 | 0.300 | 0.135 | *0.125* | **-0.034** | 0.479 |
| | | 0.092 | 0.133 | **0.012** | *0.019* | -0.034 | 0.342 |
| | | 0.229 | 0.241 | 0.059 | *0.058* | **-0.034** | 0.449 |

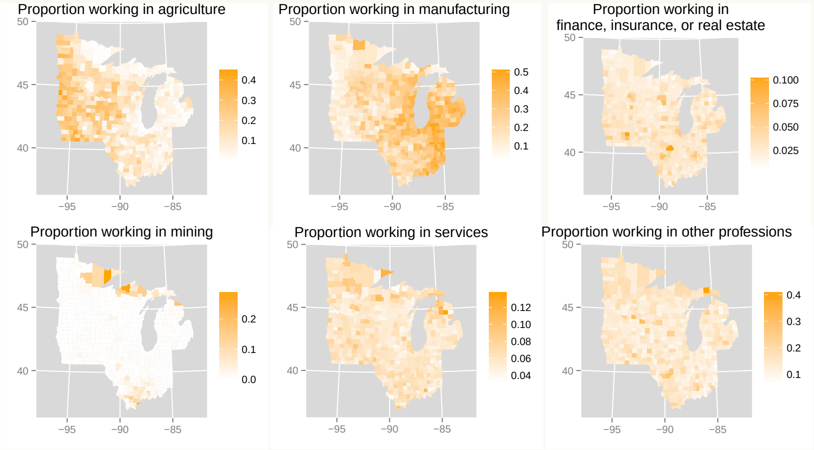
Table : Bias of $\hat{\beta}_1$ (**minimum**, *next best*).

# Section 6

## Data example: poverty rate in the upper midwest

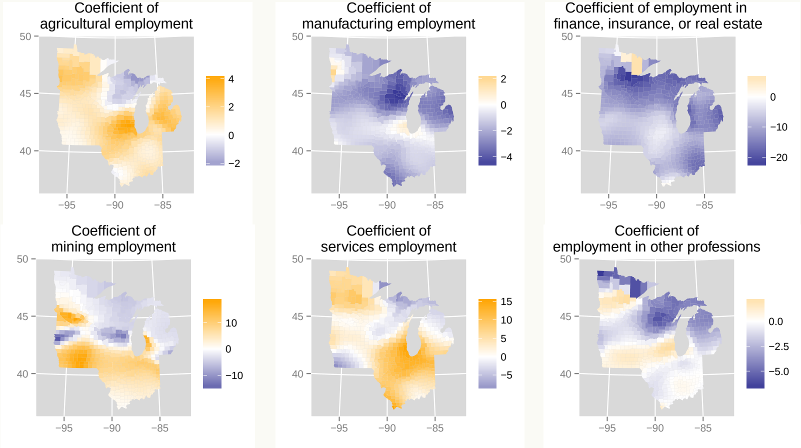
# Data example: poverty rate in the upper midwest

Revisiting the introductory example



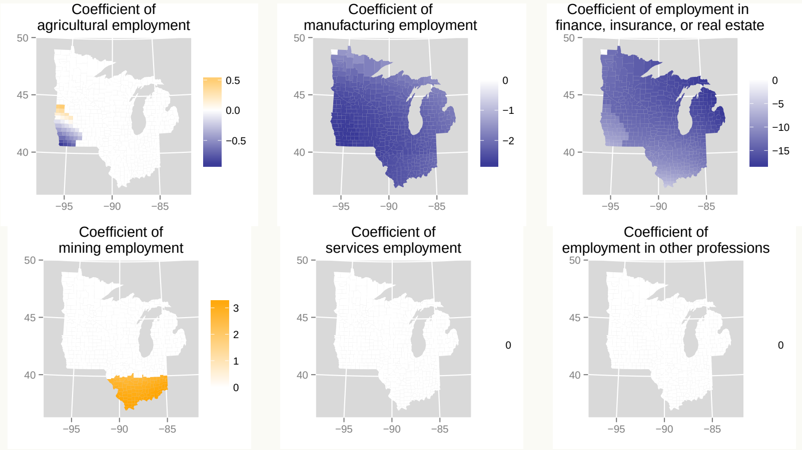
# Data example: poverty rate in the upper midwest

## Results from traditional GWR



# Data example: poverty rate in the upper midwest

Results from GWEN



# Section 7

# Future work

# Future work

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

# Acknowledgements