Are we modeling spatially varying processes or non-linear relationships?

Mehak Sachdeva

Publication: Sachdeva M., Fotheringham A. S., Li Z., and Yu H., (2022) "Are we modeling spatially varying processes or non-linear relationships?" *Geographical Analysis*

Structure

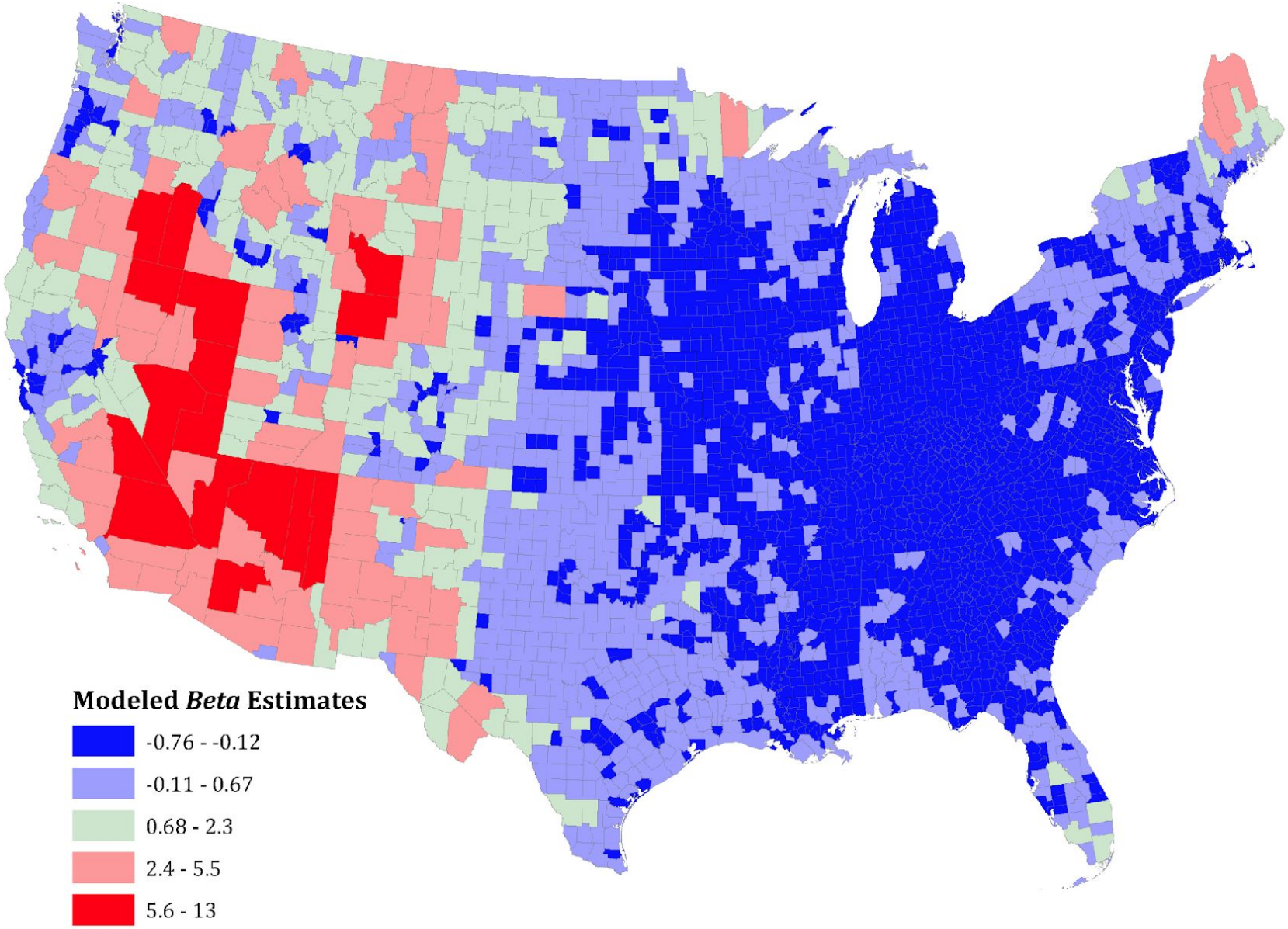
1. Problem Statement

2. Research Objective

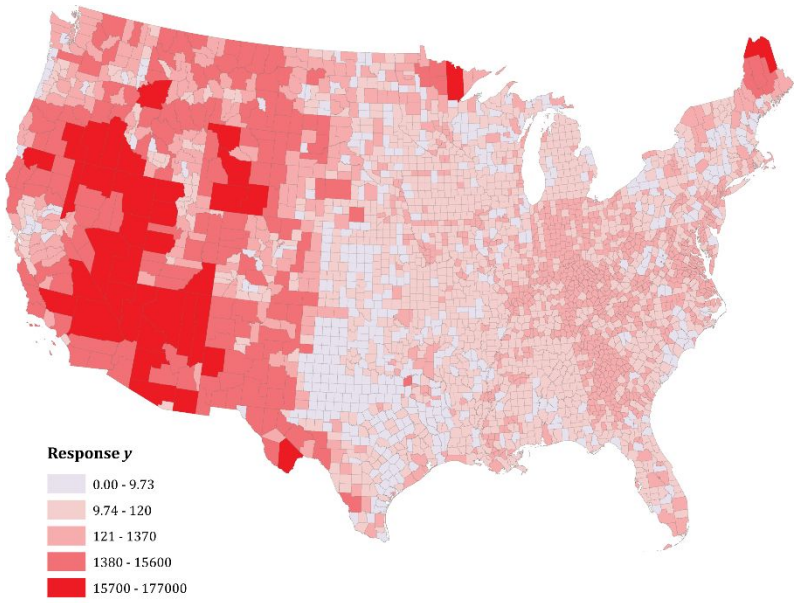
3. Research Design

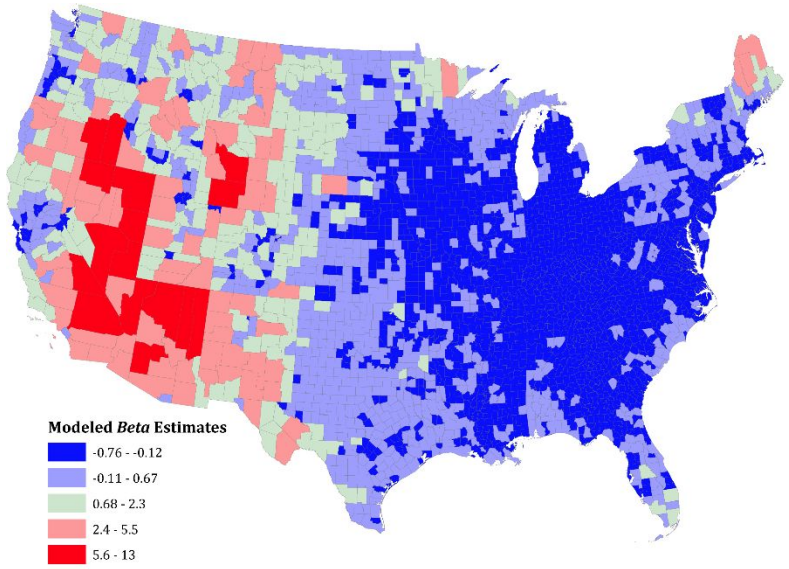
4. Intellectual Merits and Broader Impacts

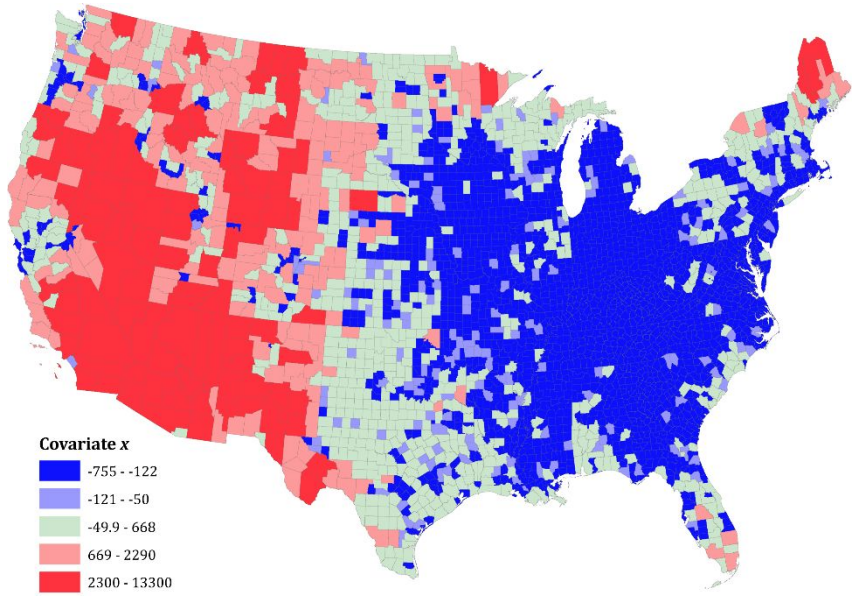
1. Problem Statement



Usual Interpretation

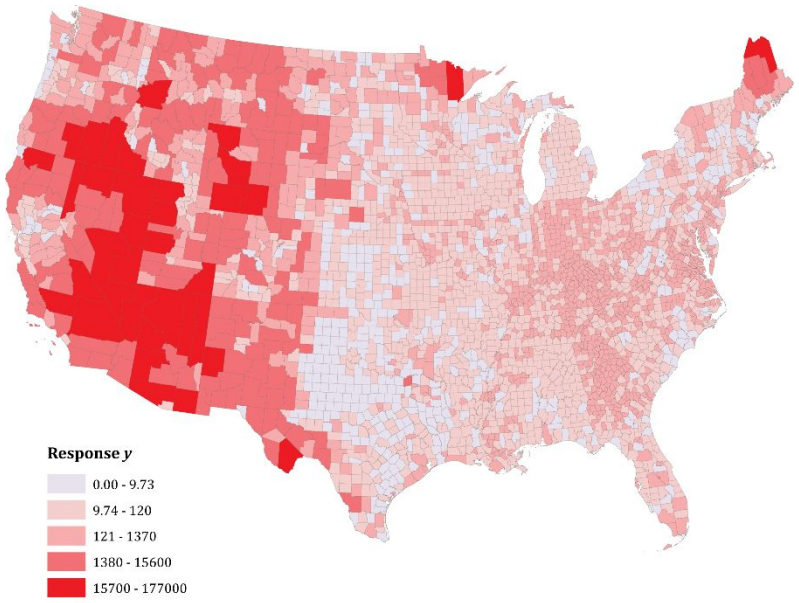


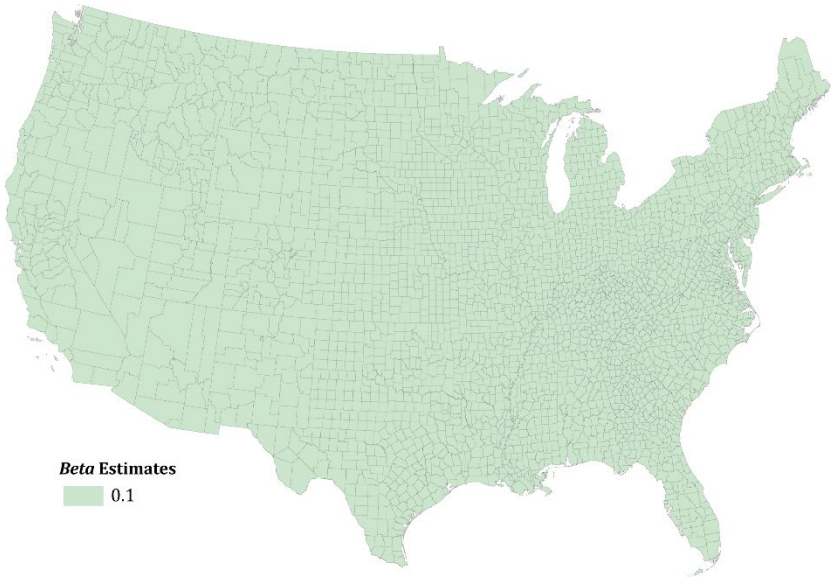


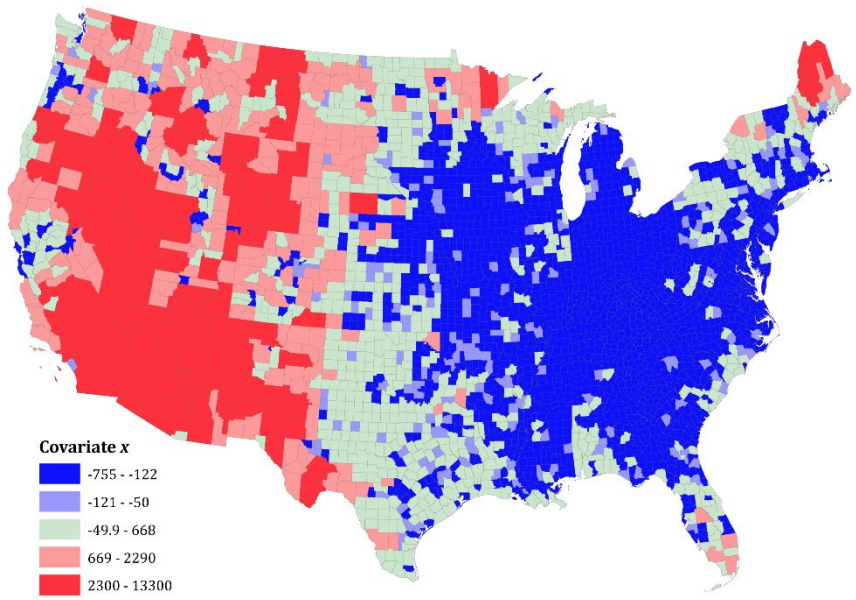


$$y\ (response) = \beta\ (process) * x\ (covariate)$$

Possible Reality



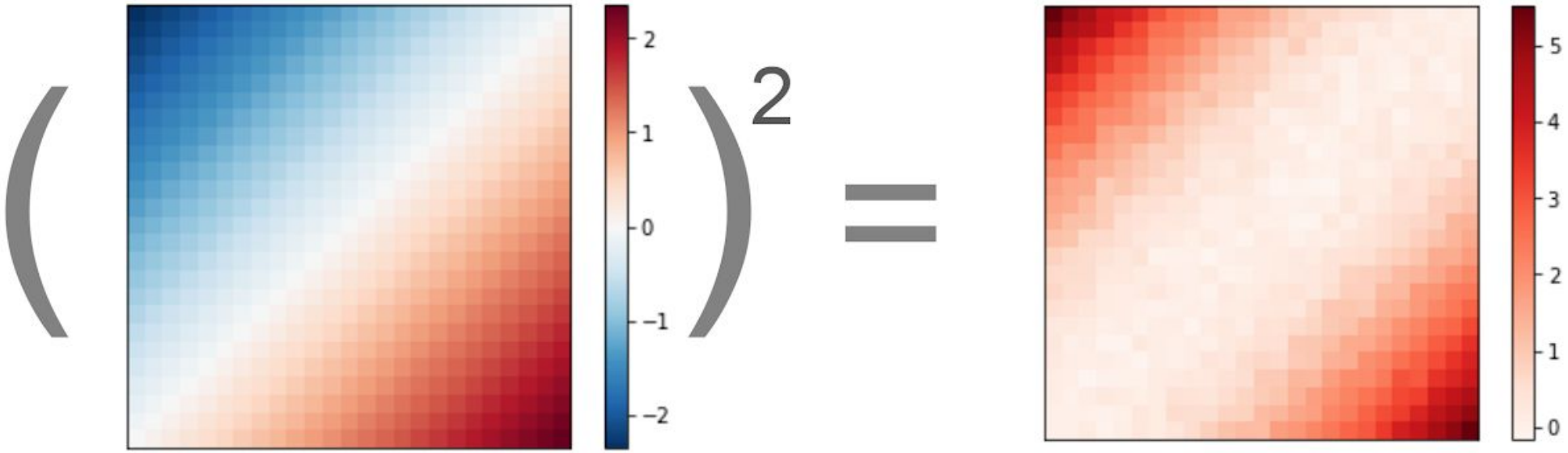




$$y\ (response) = \beta\ (process) * x^2\ (covariate)$$

Problem Statement

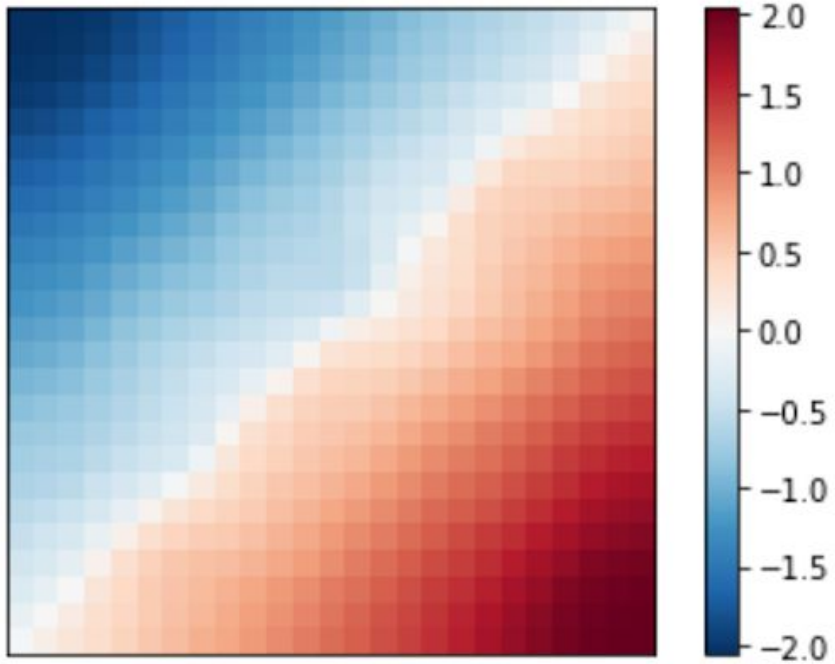
Consider a situation such that:



(a). Covariate x

(b). Response variable y

Problem Statement



Parameter estimates from MGWR

- Without examining for the possibility of nonlinearity, the spatial variation in this surface would then be incorrectly interpreted to represent process spatial non-stationarity.

- How do we differentiate such situations from presence of non-stationarity in processes?

Background

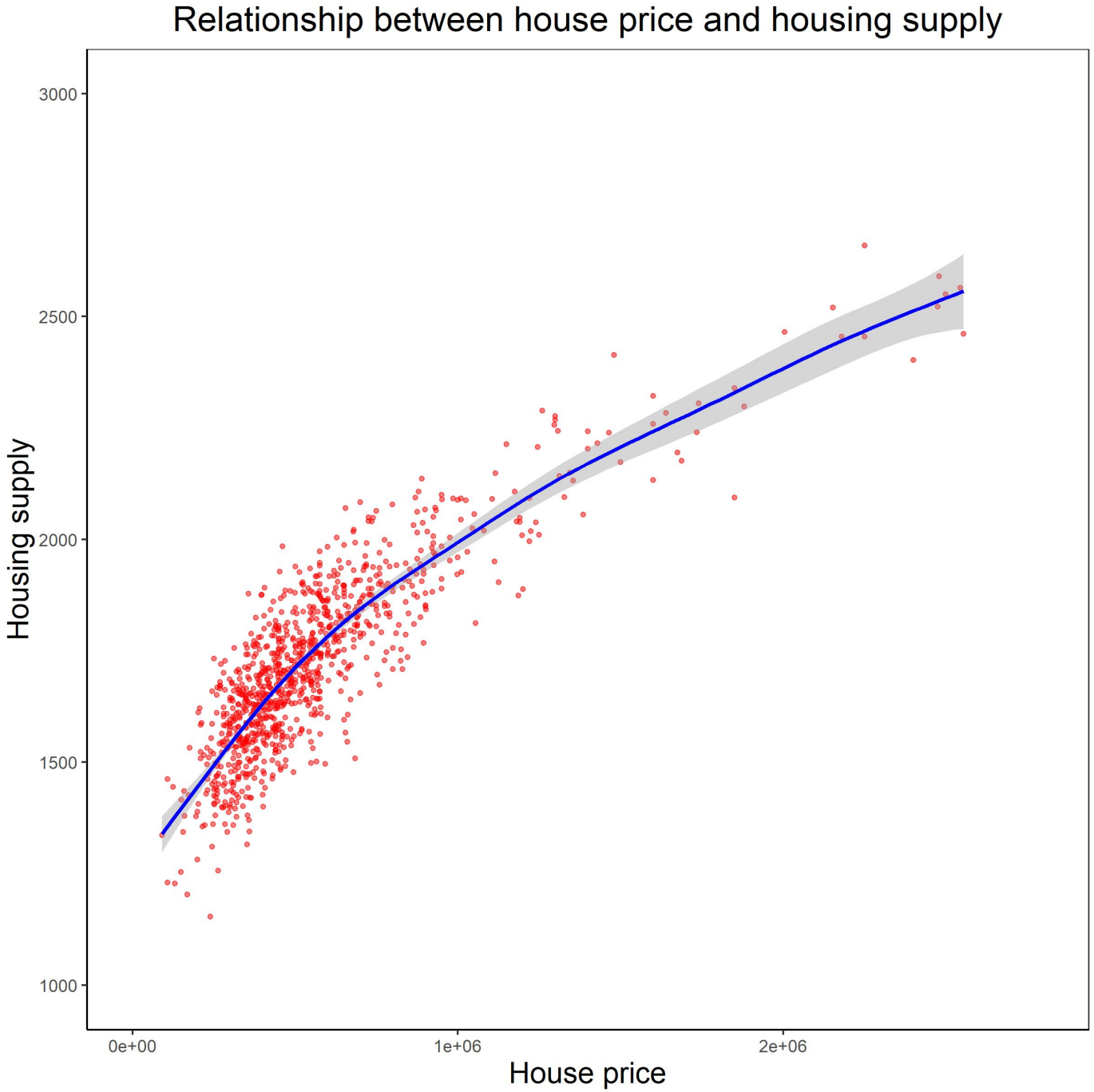
Why is this problem important?

$$housing\ supply = \beta_{0(u_i,v_i)} + \beta_{1(u_i,v_i)}(house\ price)$$

(Note: Hypothetical example)

Background

True relationship

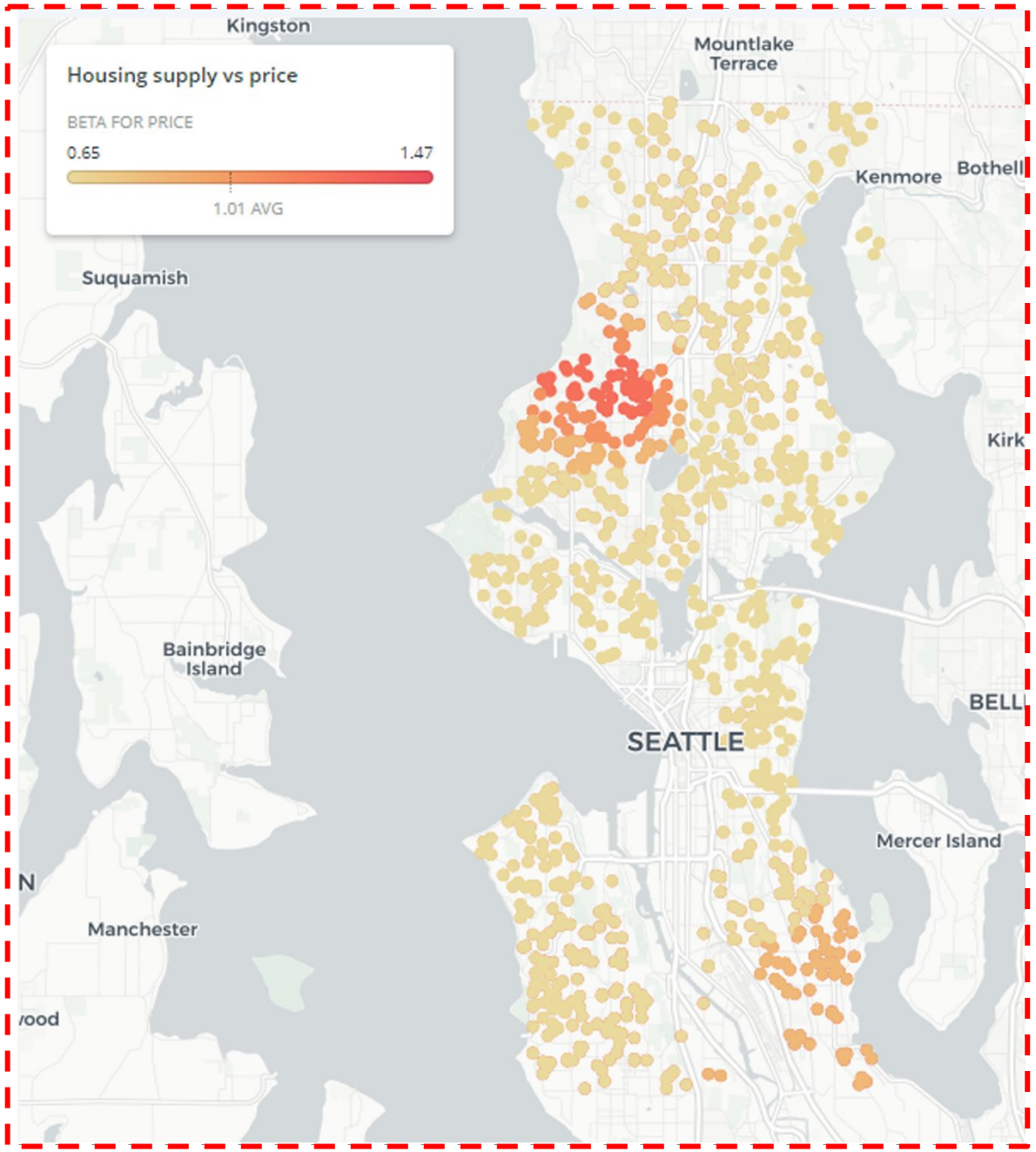


(Note: Hypothetical example)

Background

Estimated relationship

$$housing\ supply = \beta_{0(u_i,v_i)} + \beta_{1(u_i,v_i)}(\sqrt{house\ price})$$



(Note: Hypothetical example)

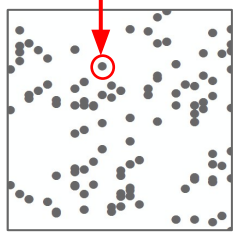
So, we need to separate these two scenarios

Modeling Processes

Process Spatial Nonstationarity

Local Models e.g. MGWR

$$y_i = \sum_j \beta_{ij}(u_i, v_i) X_{ij} + \varepsilon_i$$

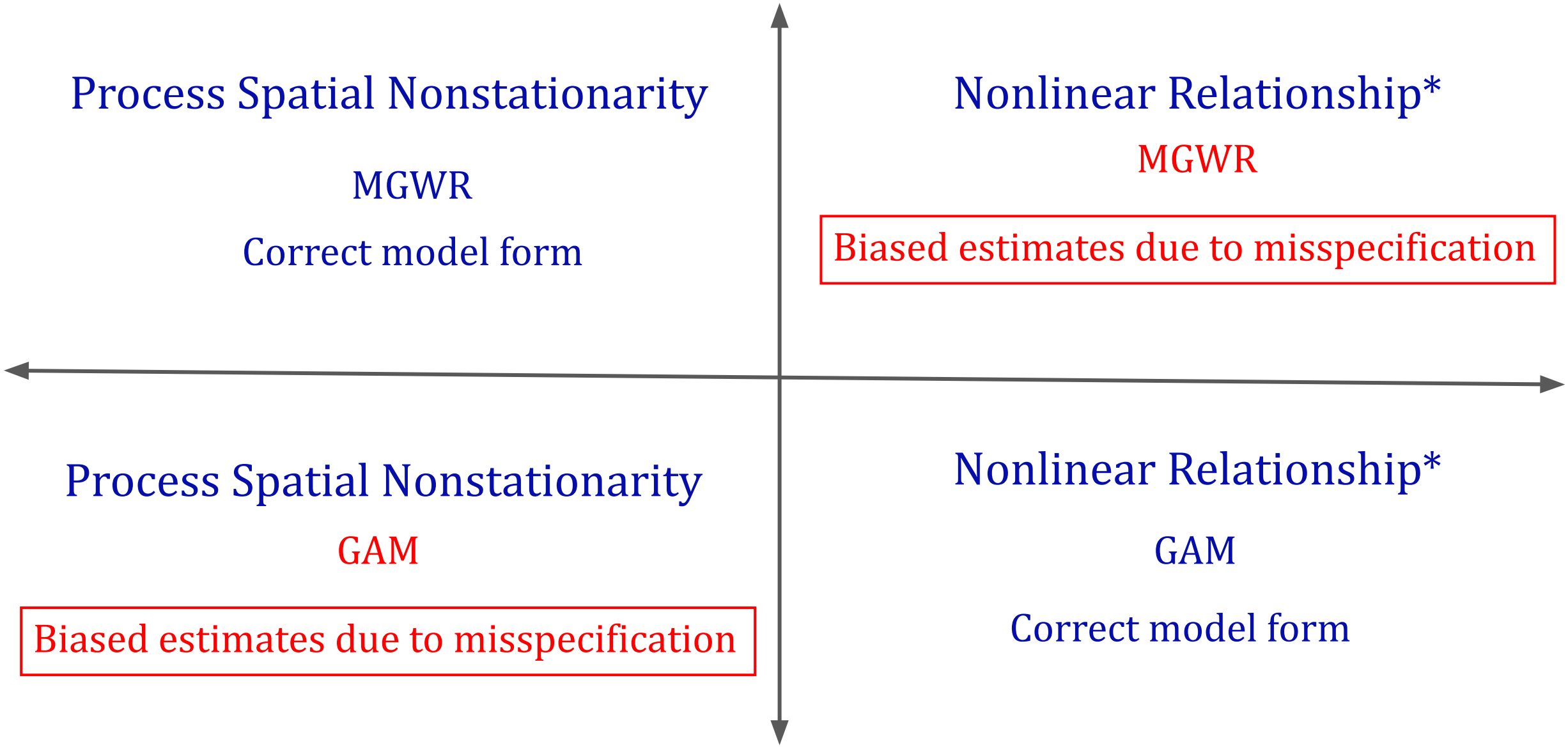


Nonlinear Relationships

Models e.g. GAM

$$y_i = \sum_{j=1}^{p} f_j(x_{ij})$$

Otherwise we make incorrect inferences about processes



* Assuming the covariate is spatially varying

Hence:

1. Is there an easy way to **detect when spatially varying parameter estimates are masking nonlinear relationships?**

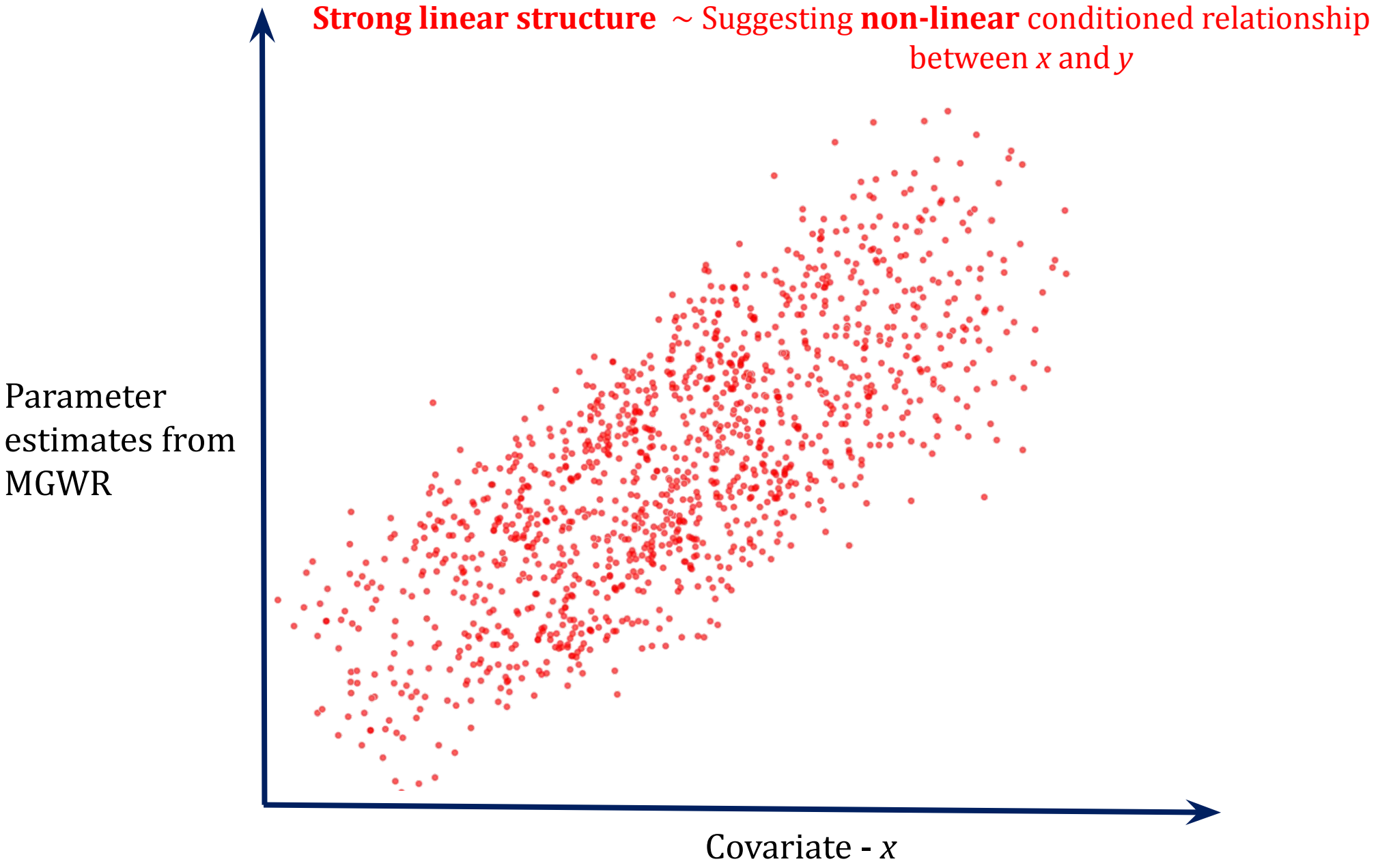
2. Conversely, is there an easy way to **detect when estimated nonlinear relationships are masking process nonstationarity?**

2. Research Objectives

- **Describe a simple test** to check whether parameter estimates generated by local models are manifestations of nonlinear relationships rather than spatially varying processes

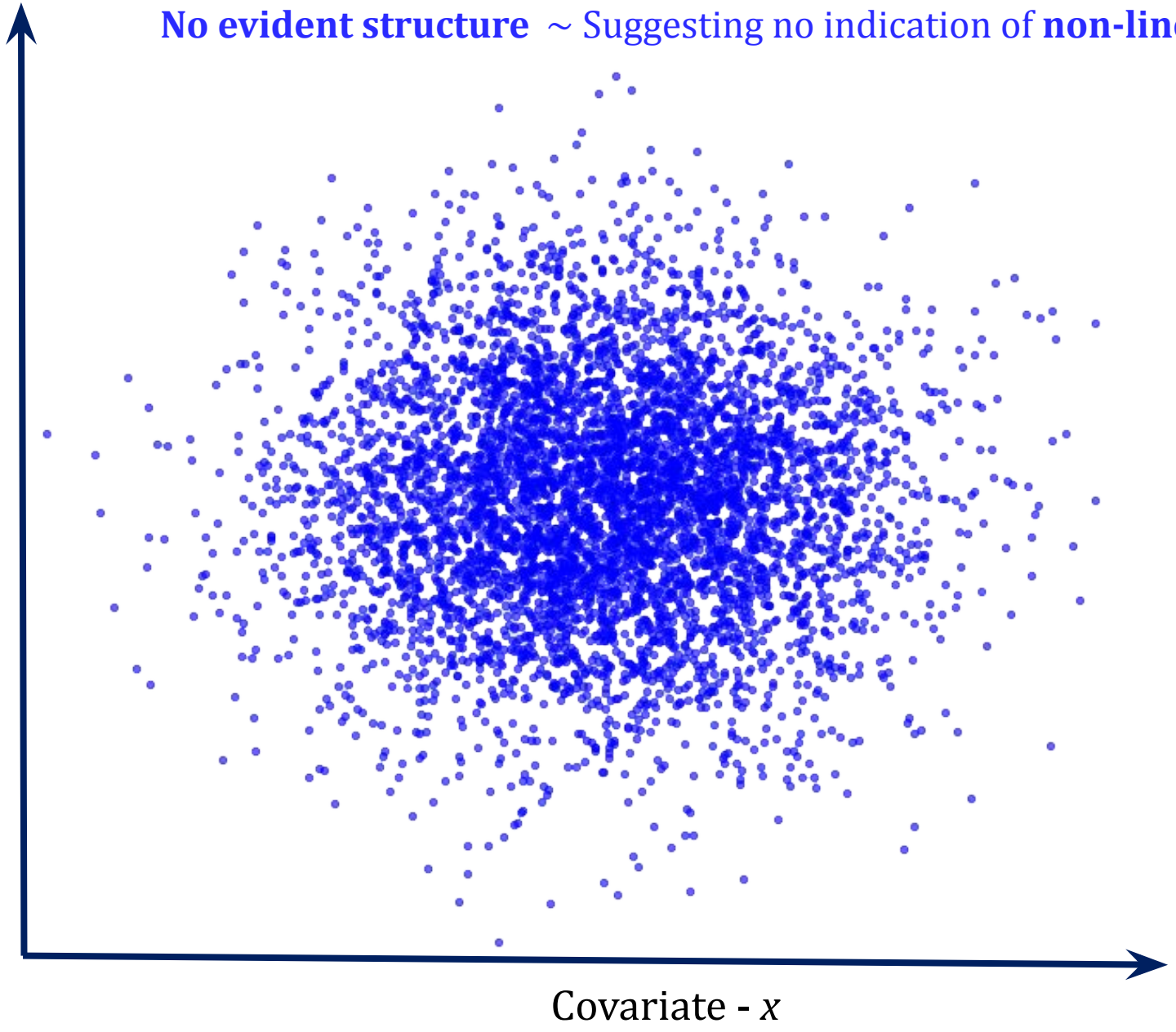
- **Demonstrate feasibility of such a test** in simulation studies and empirical research with real-world noisy data

Diagnostic test described

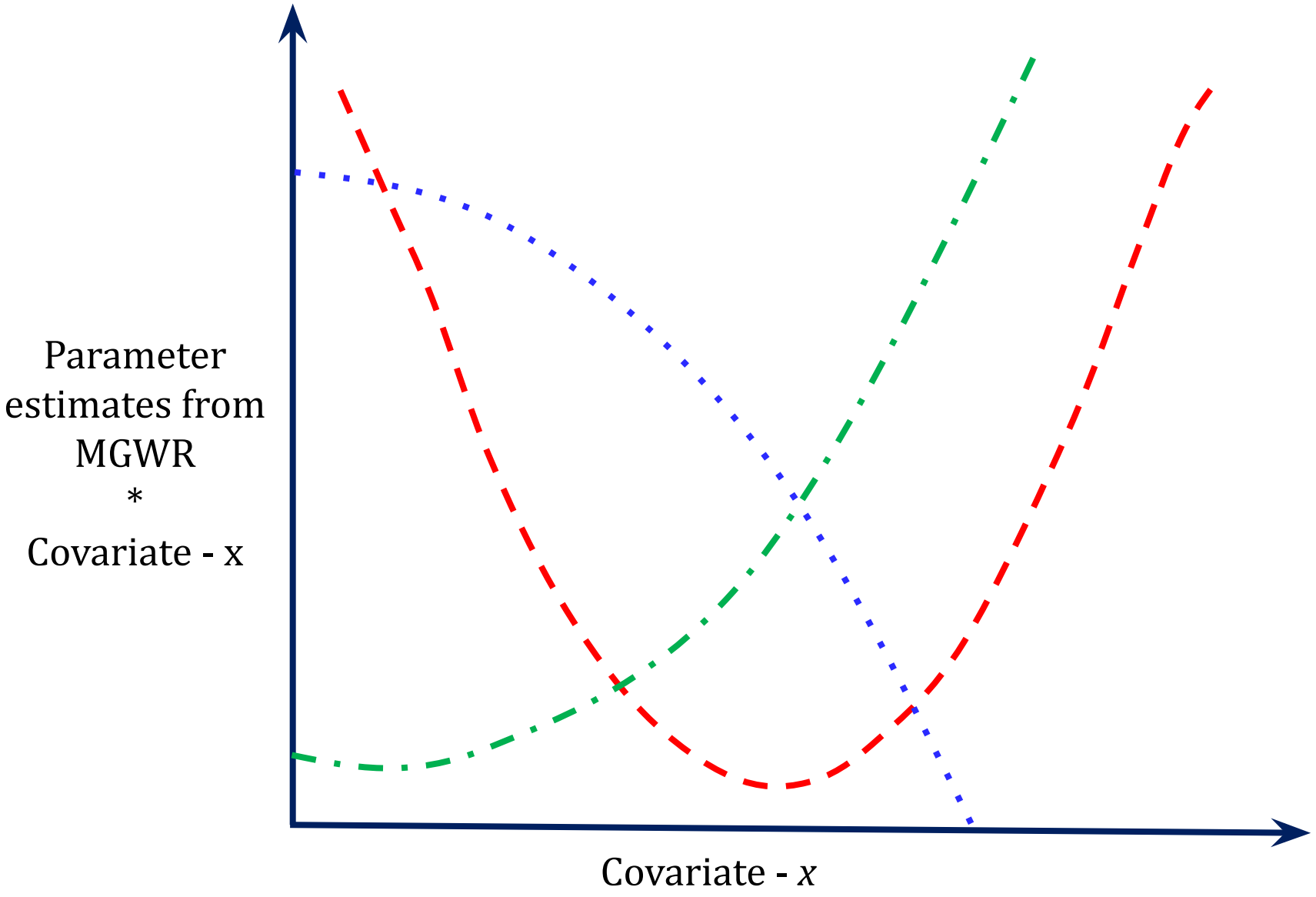


Diagnostic test described

No evident structure ~ Suggesting no indication of **non-linearity**

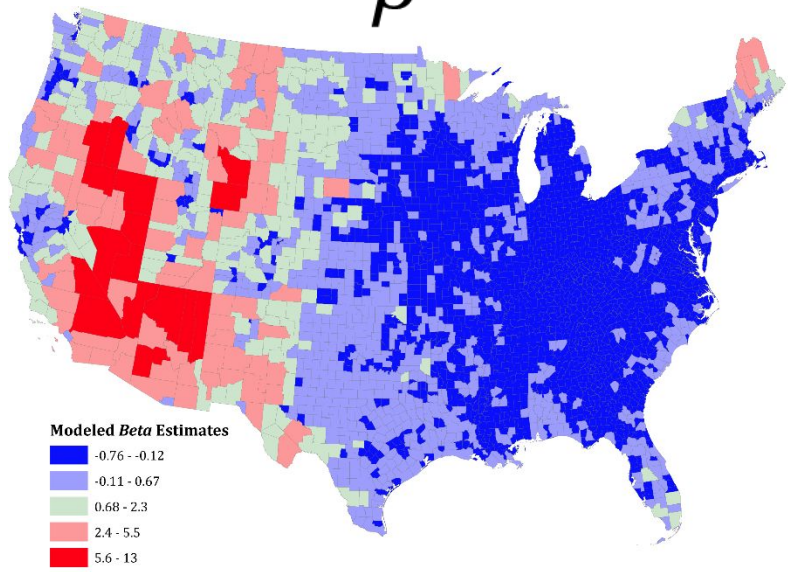


Plot for further evidence on type of non-linearity

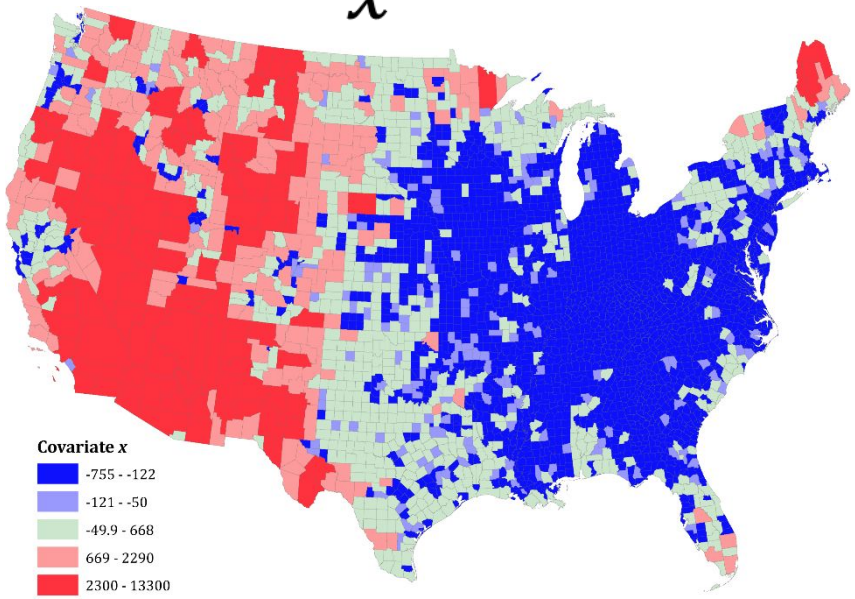


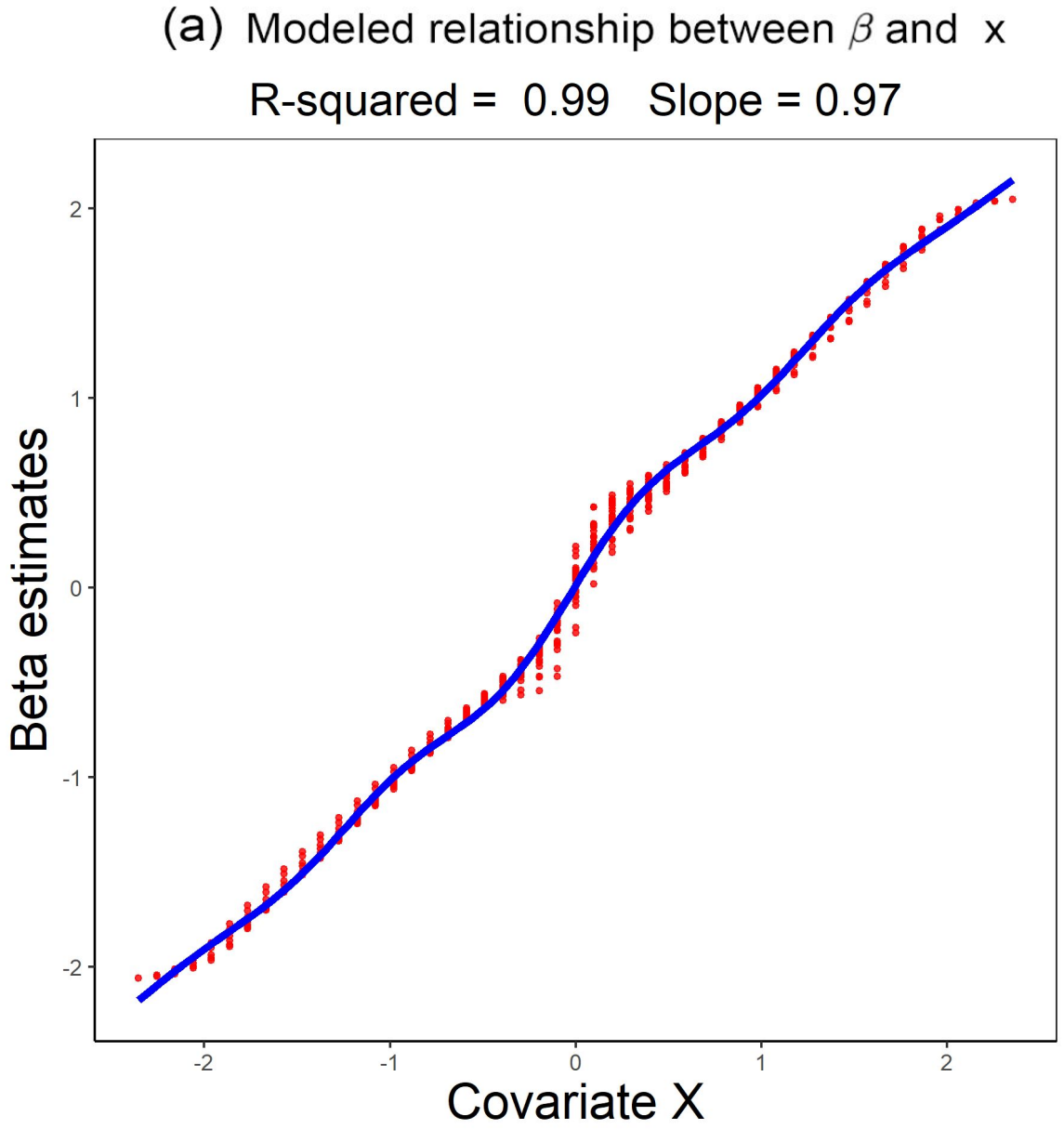
Revisiting the example:

β

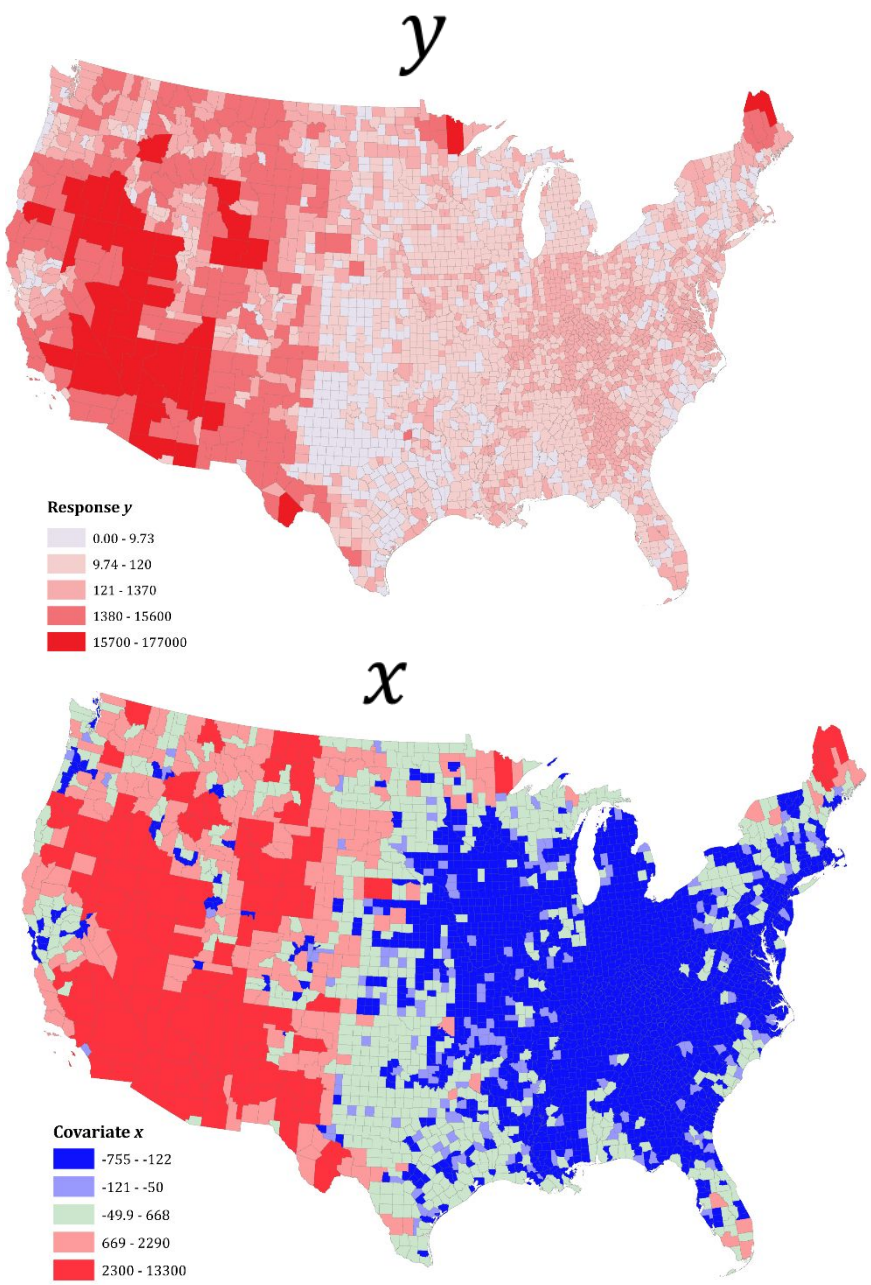


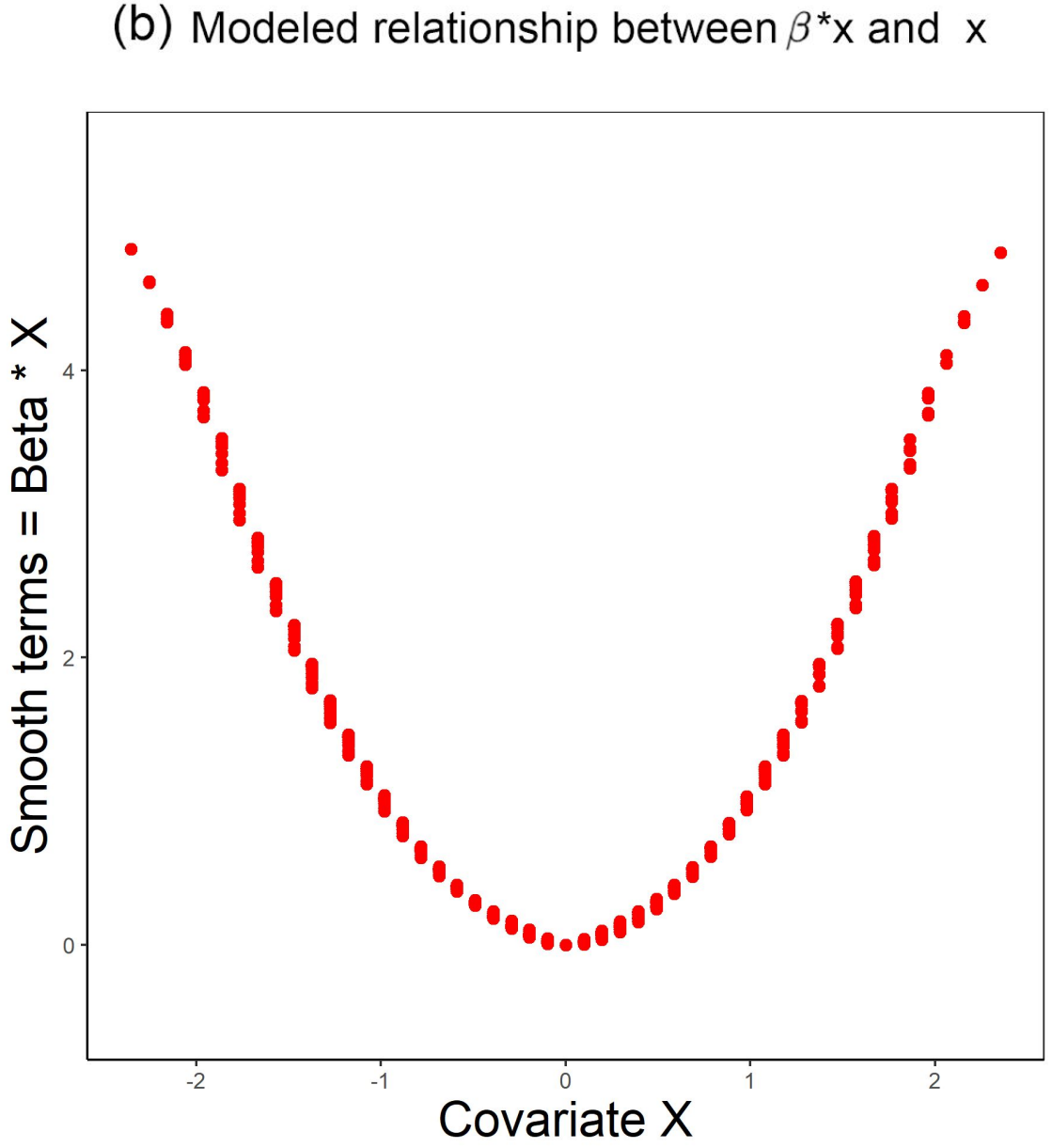
x





Revisiting the example:





3. Research Design

Check Feasibility of Diagnostic Tests in Two Extreme Scenarios

Experiment I

Processes are **spatially nonstationary** and **independent** of the covariates

$$\boldsymbol{y = \beta_1 * x_1 + \beta_2 * x_2 + \epsilon}$$

β_1 *and* x_1 *are independent*
β_2 *and* x_2 *are independent*

Experiment II

Processes are **spatially stationary** and **derived** from the covariates

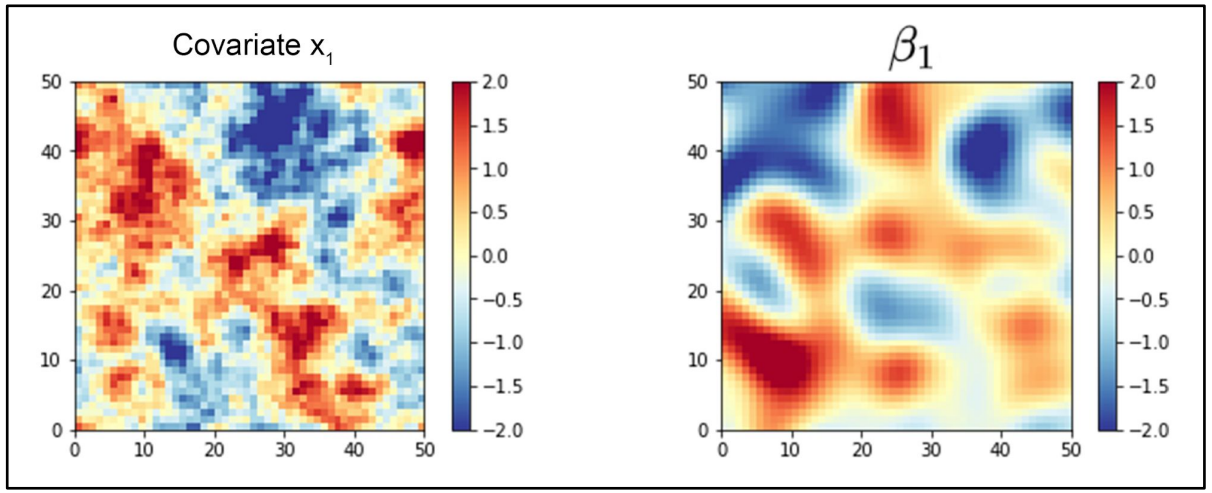
$$y = \beta_1 * x_1 + \beta_2 * x_2 + \epsilon$$

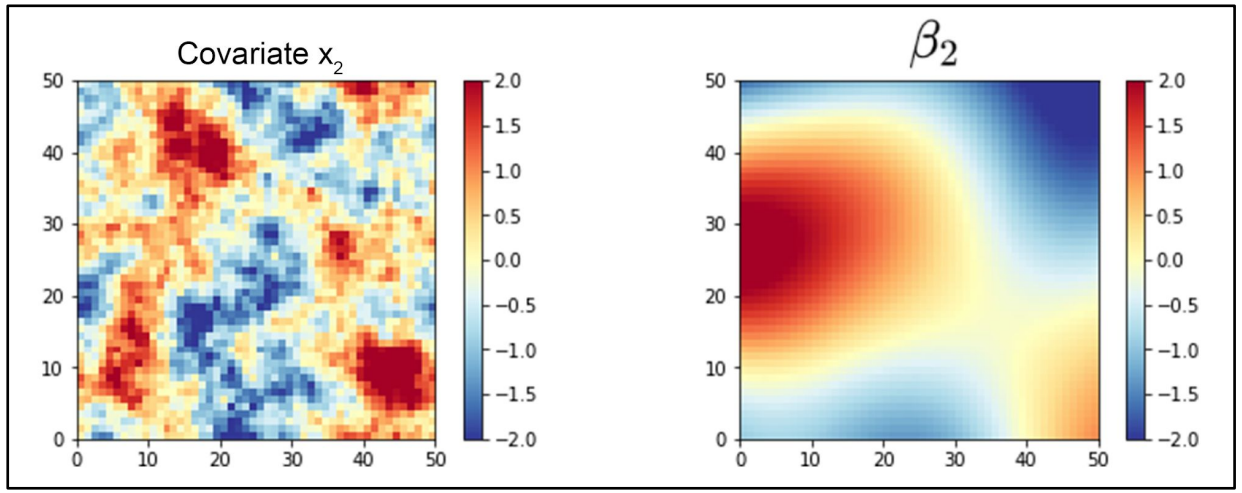
$$\beta_1 = k_1 * x_1 \; and \; \beta_2 = k_2 * \; x_2$$

$$\boldsymbol{y = k_1 * x_1^2 + k_2 * x_2^2 + \epsilon}$$

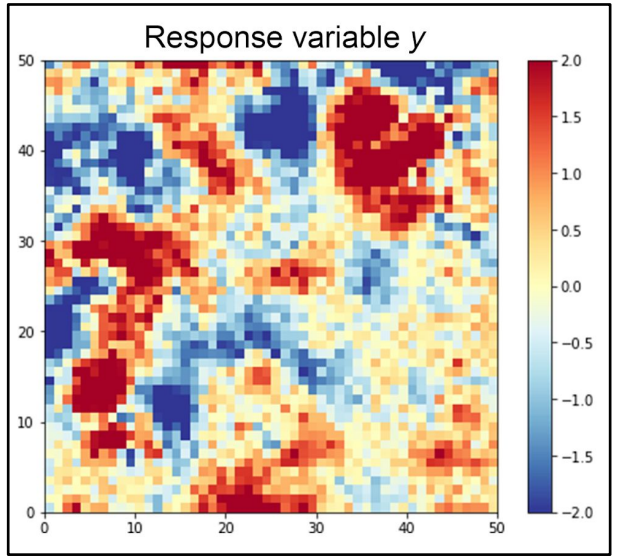
k_1*and* k_2 *are near constant terms*

Experiment I – Non-stationarity in processes

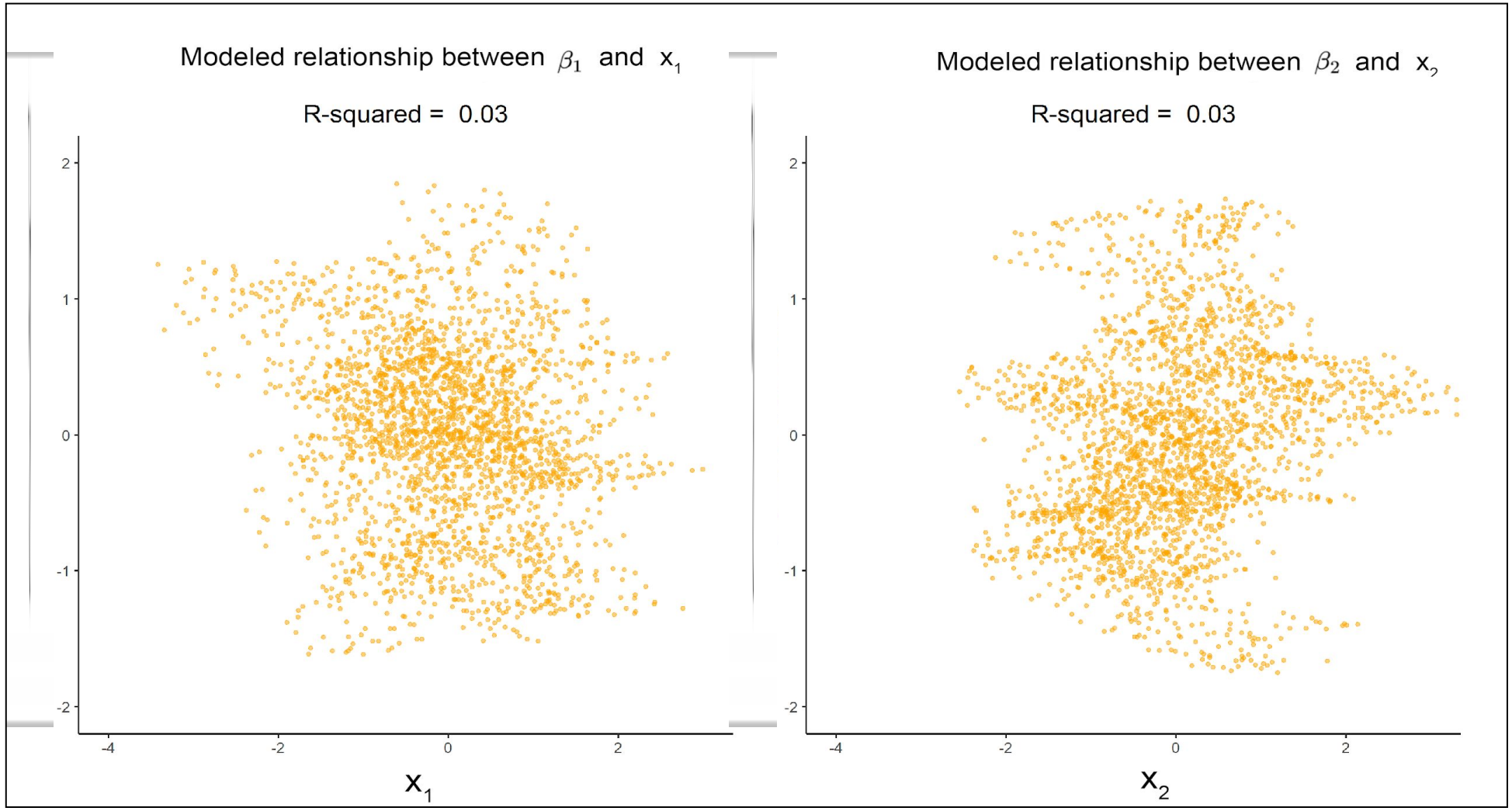




$$y = \beta_1 * x_1 + \beta_2 * x_2 + \epsilon$$

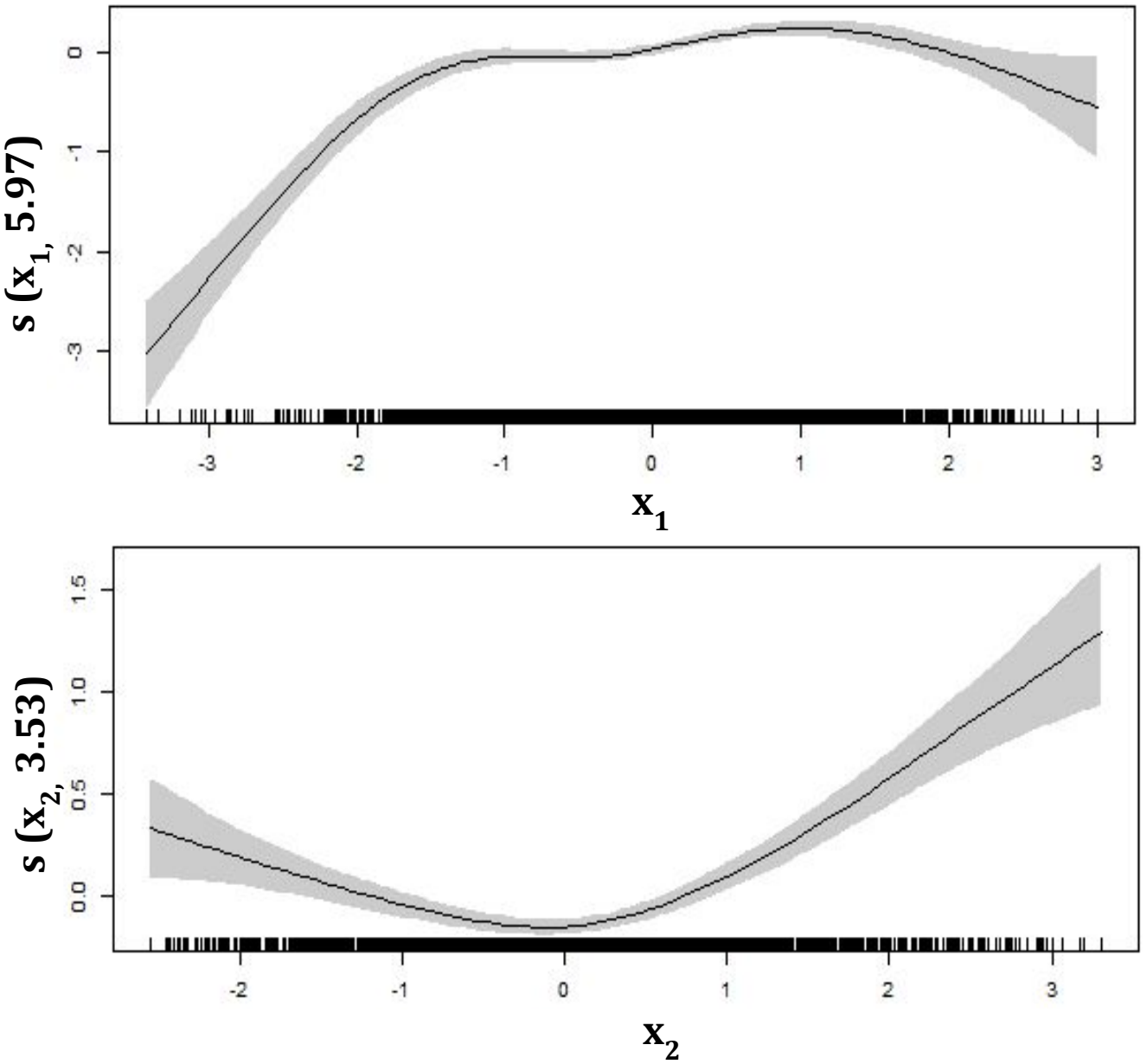


Experiment I – Diagnostic test: Detecting nonlinearity as a cause of spatially varying parameter estimates



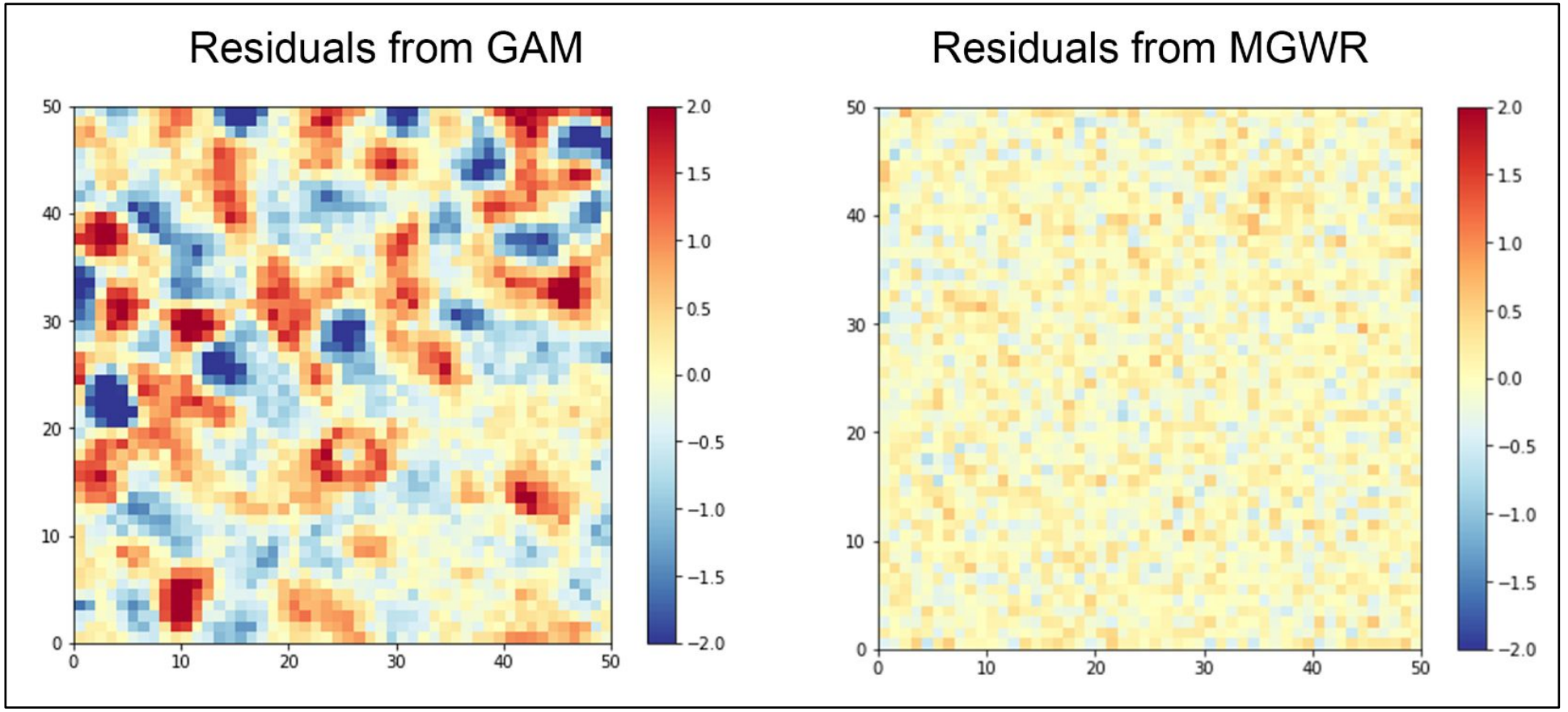
Nonlinearity Not Detected

Experiment I – GAM result plots



Experiment I – Diagnostic test: Detecting spatial nonstationarity when nonlinearity is assumed

Residuals from GAM and MGWR

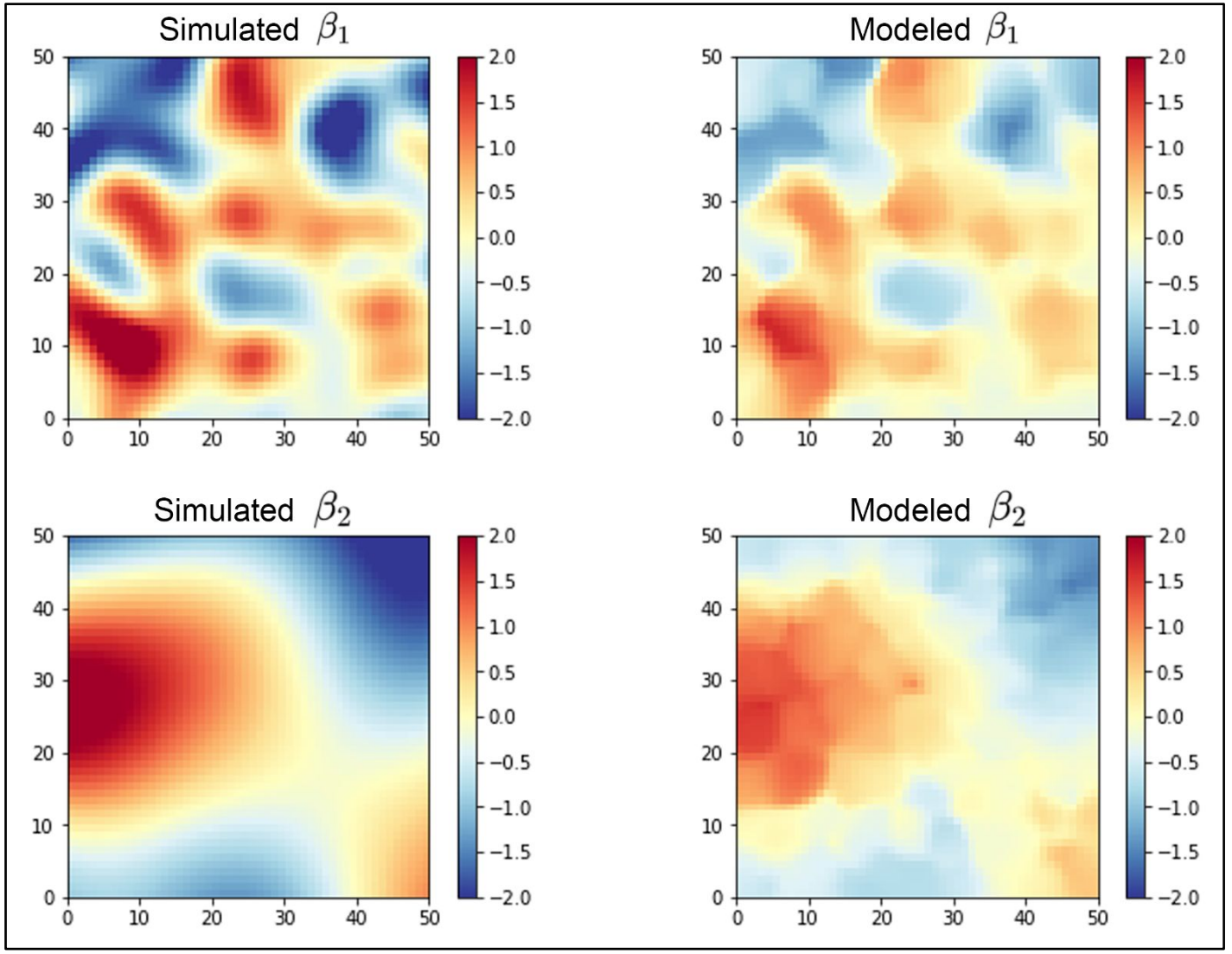


Moran's I value: +0.73
(p-value: 0.001)

Process Nonstationarity Detected

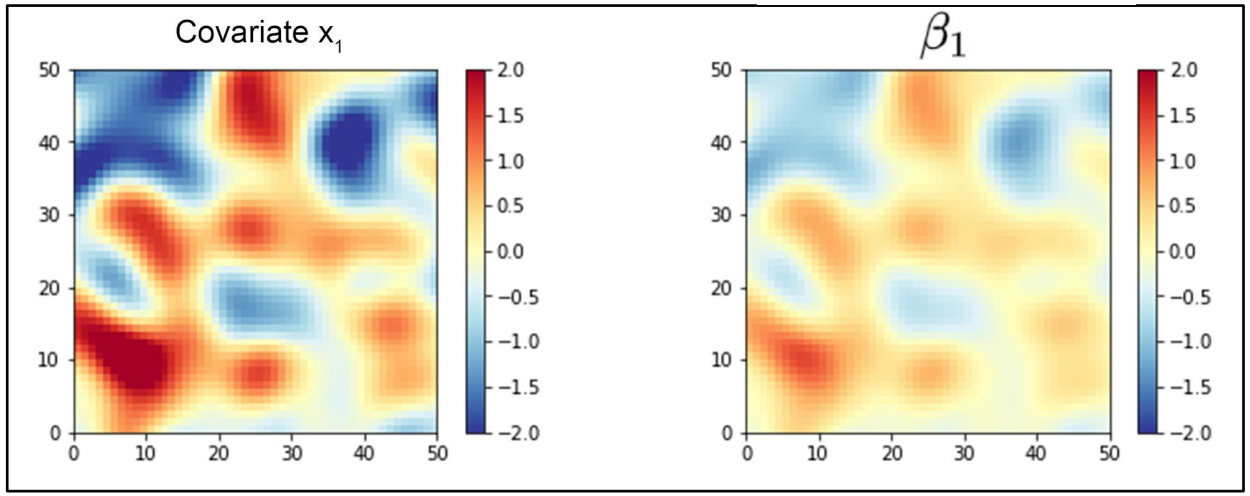
Experiment I - Further evidence of non-stationarity in processes

Simulated vs Modeled parameter estimates from MGWR

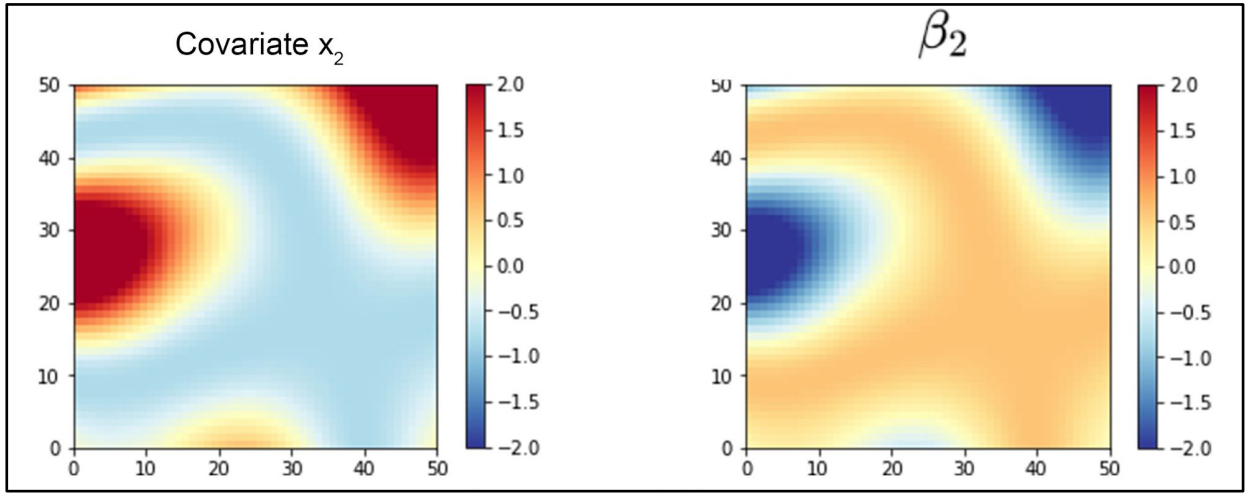


Experiment II – Complete non-linearity in relationships

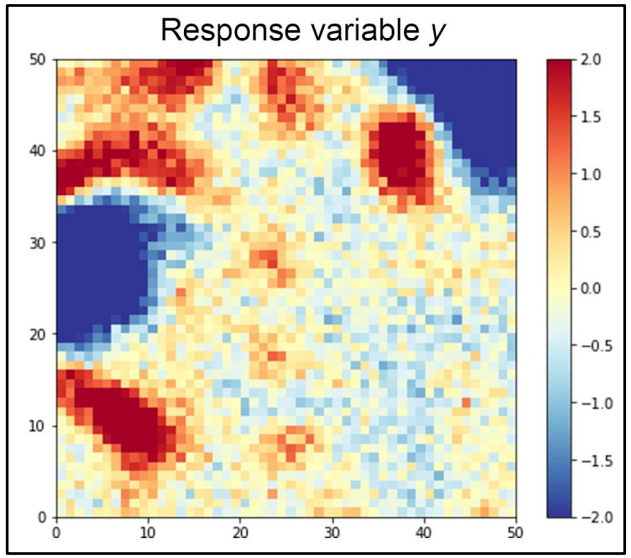
$$\beta_1 = 0.5 * x_1$$



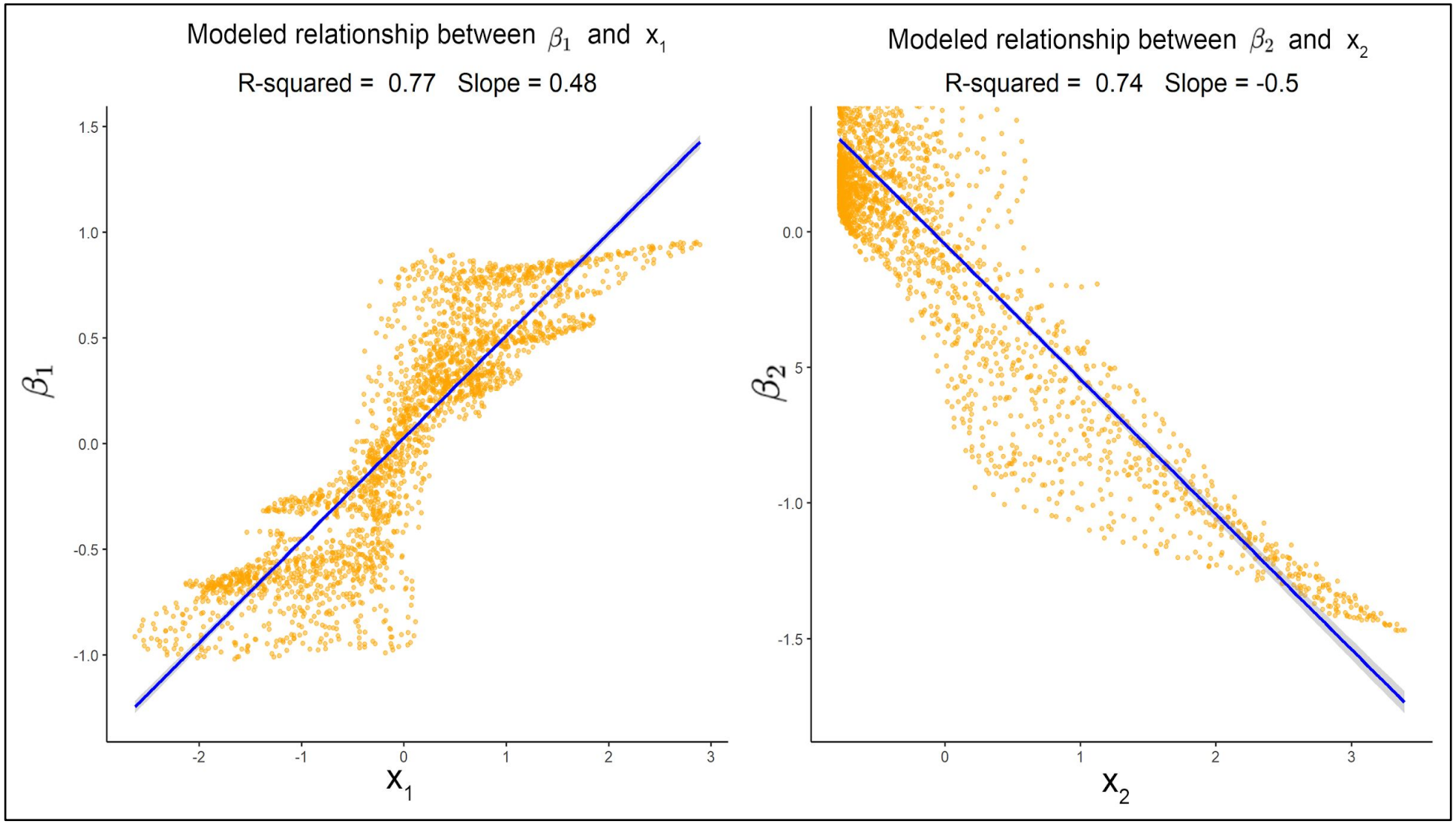
$$\beta_2 = -0.8 * x_2$$



$$y = 0.5 * x_1^2 - 0.8 * x_2{}^2 + \epsilon$$

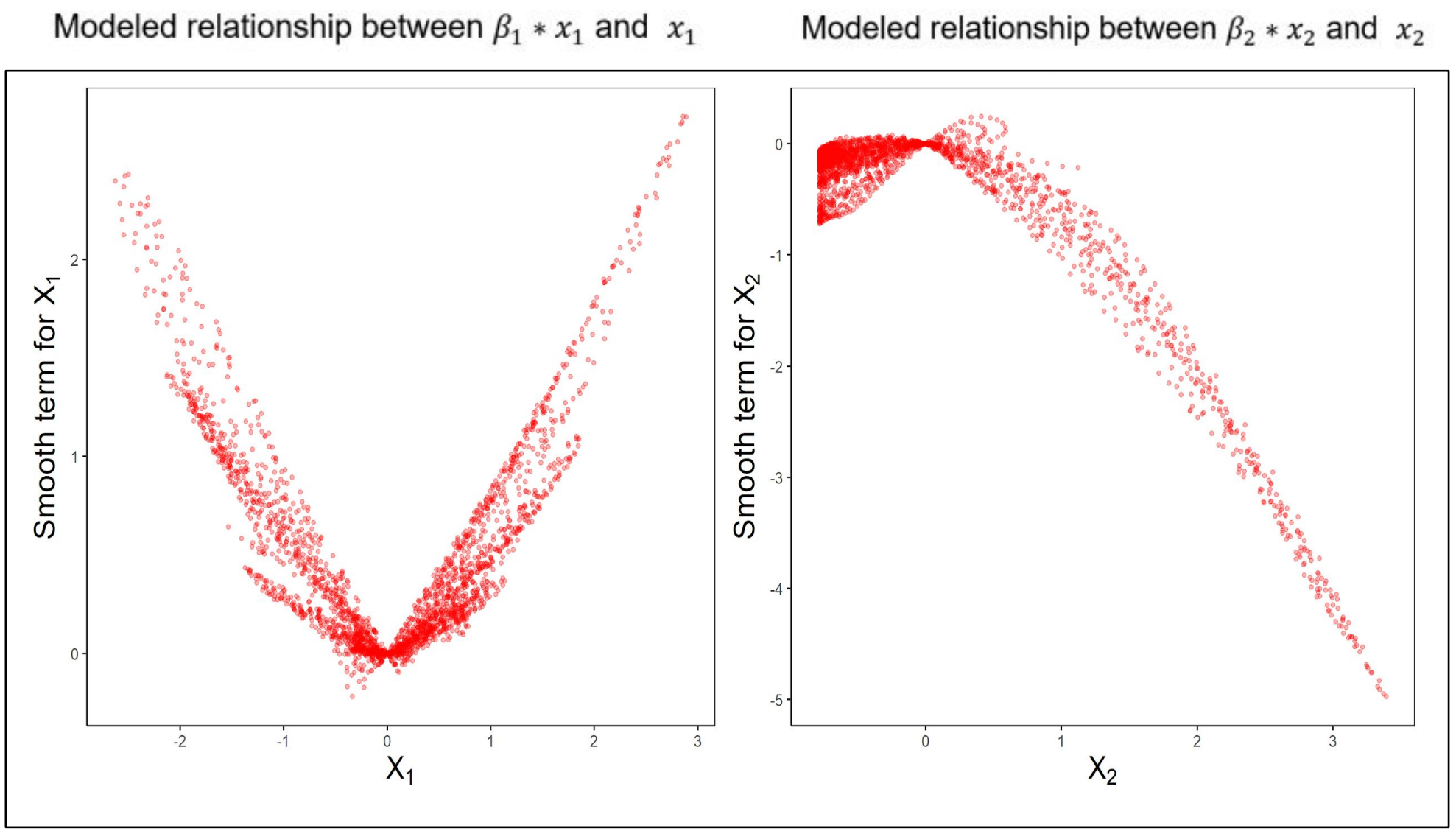


Experiment II – Diagnostic test: Detecting nonlinearity when spatial nonstationarity is assumed



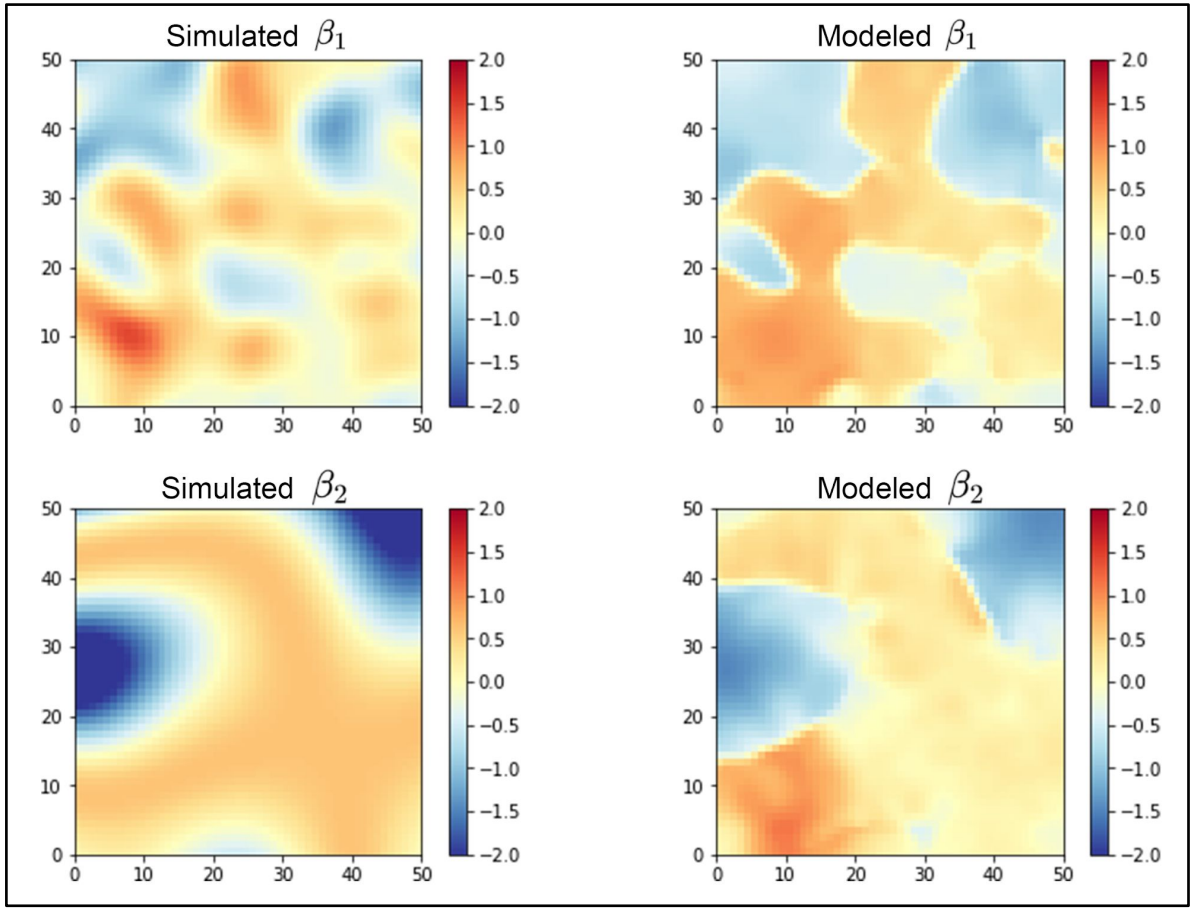
Nonlinearity Detected

Experiment II: Plot for further evidence on type of non-linearity

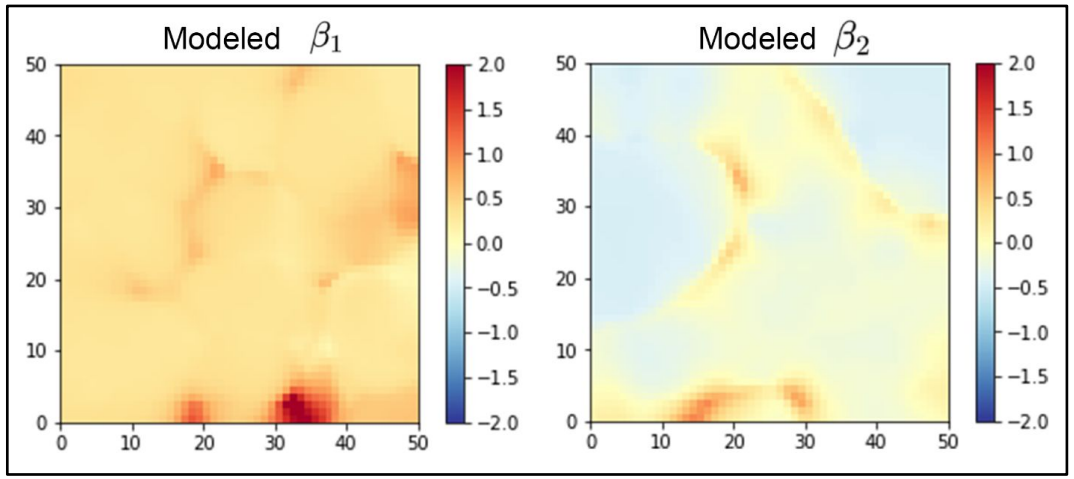


Experiment II: Parameter estimates from MGWR

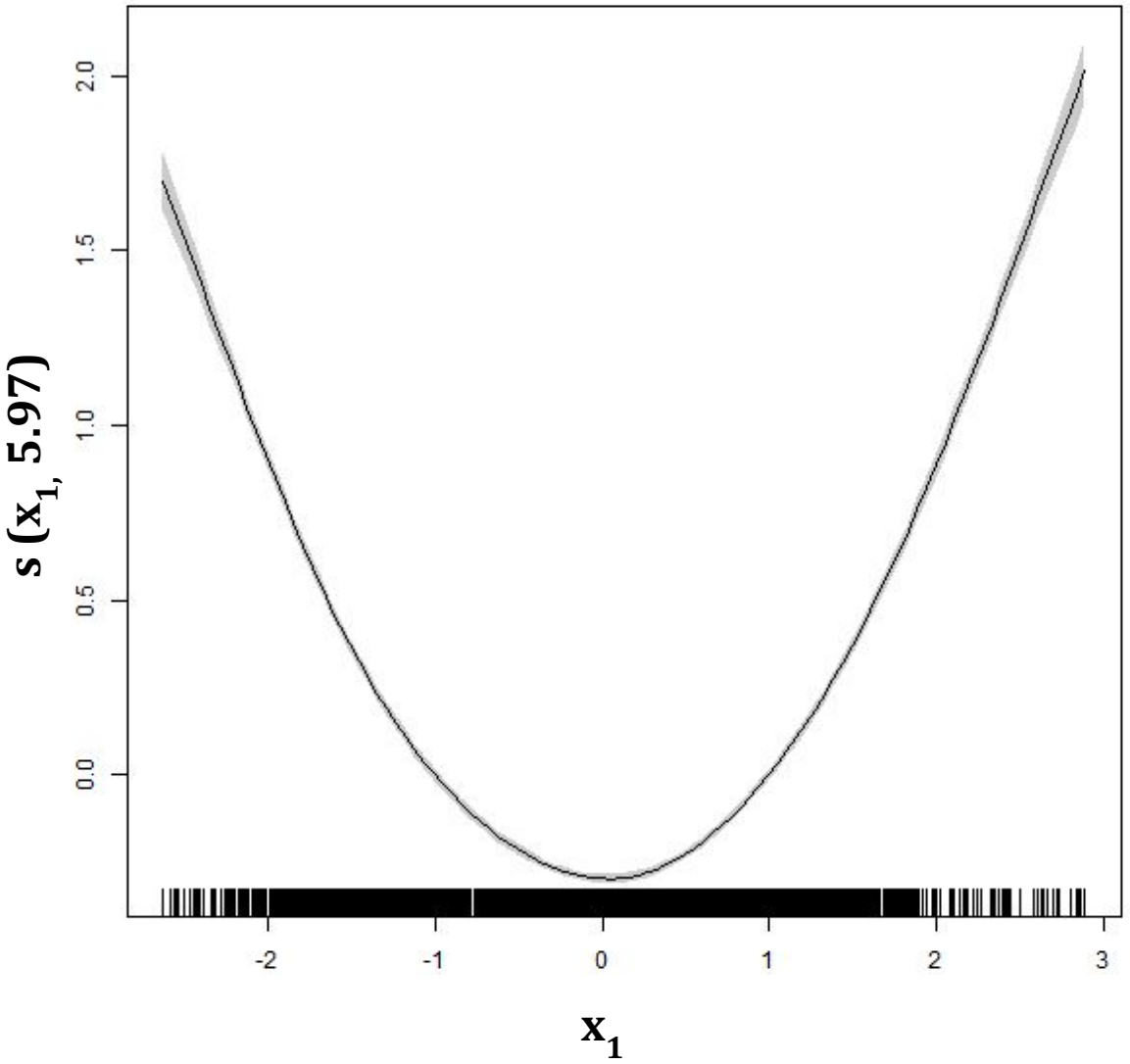
Using x_1 and x_2 as predictors

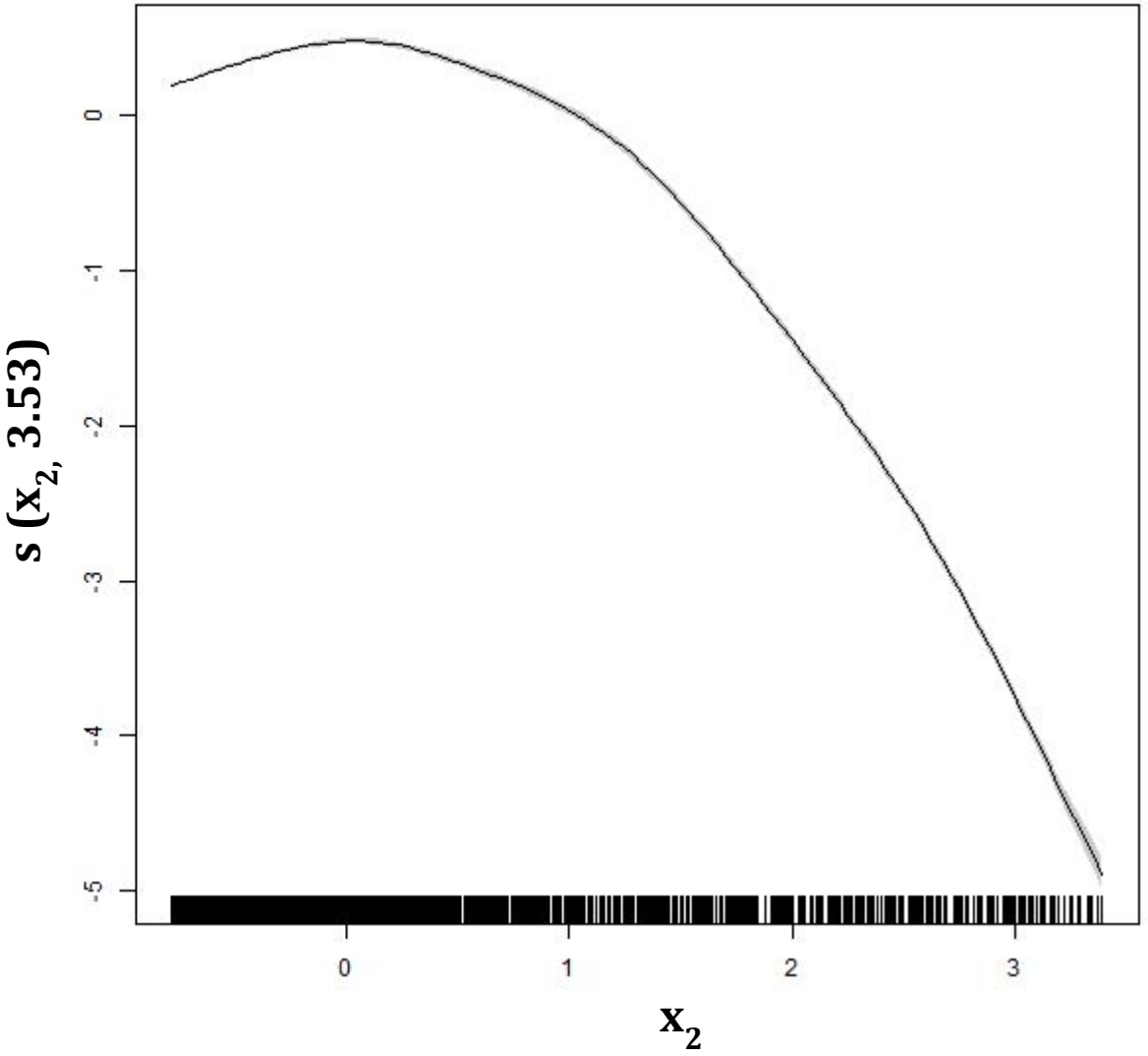


Using x_1^2 and x_2^2 as predictors



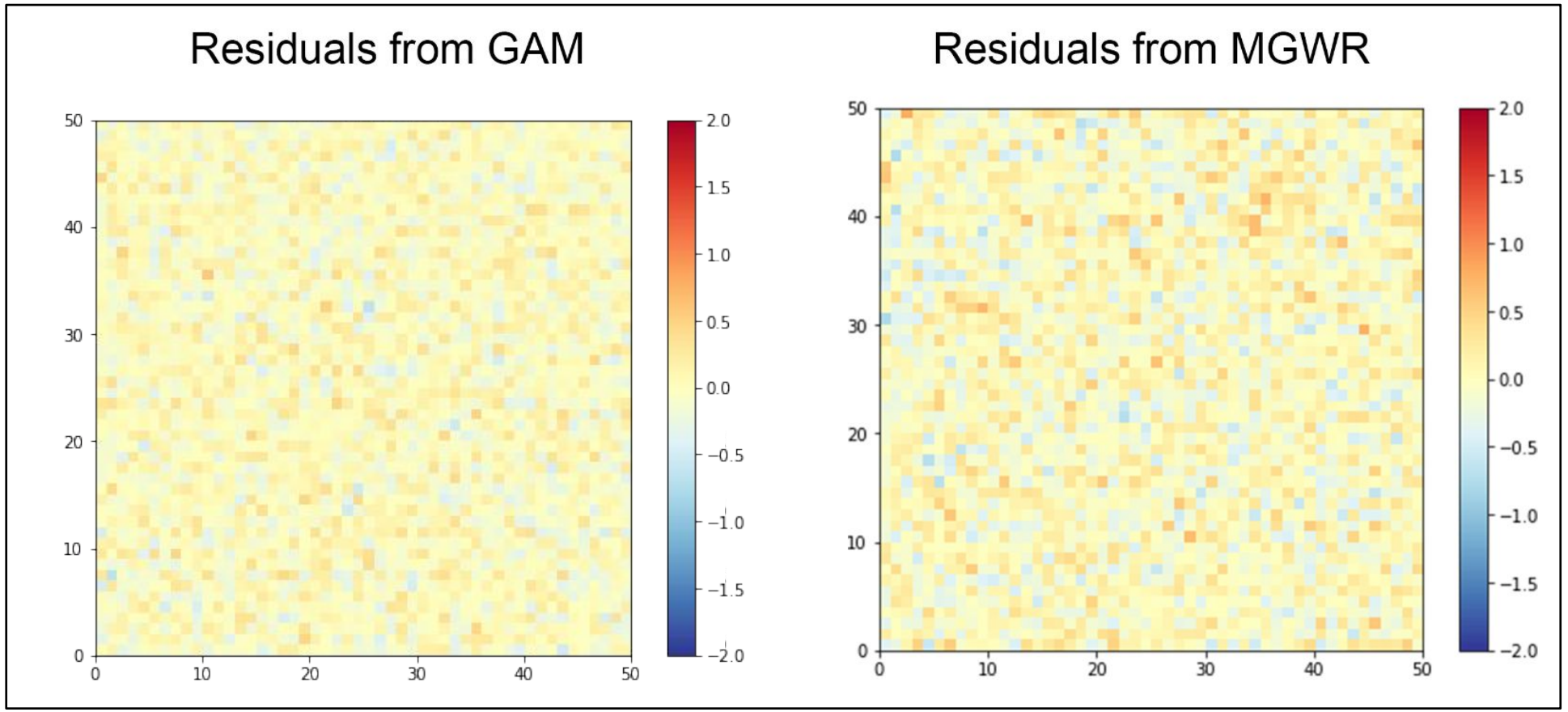
Experiment II – GAM result plots





Experiment II – Diagnostic test: Detecting spatial nonstationarity when nonlinearity is assumed

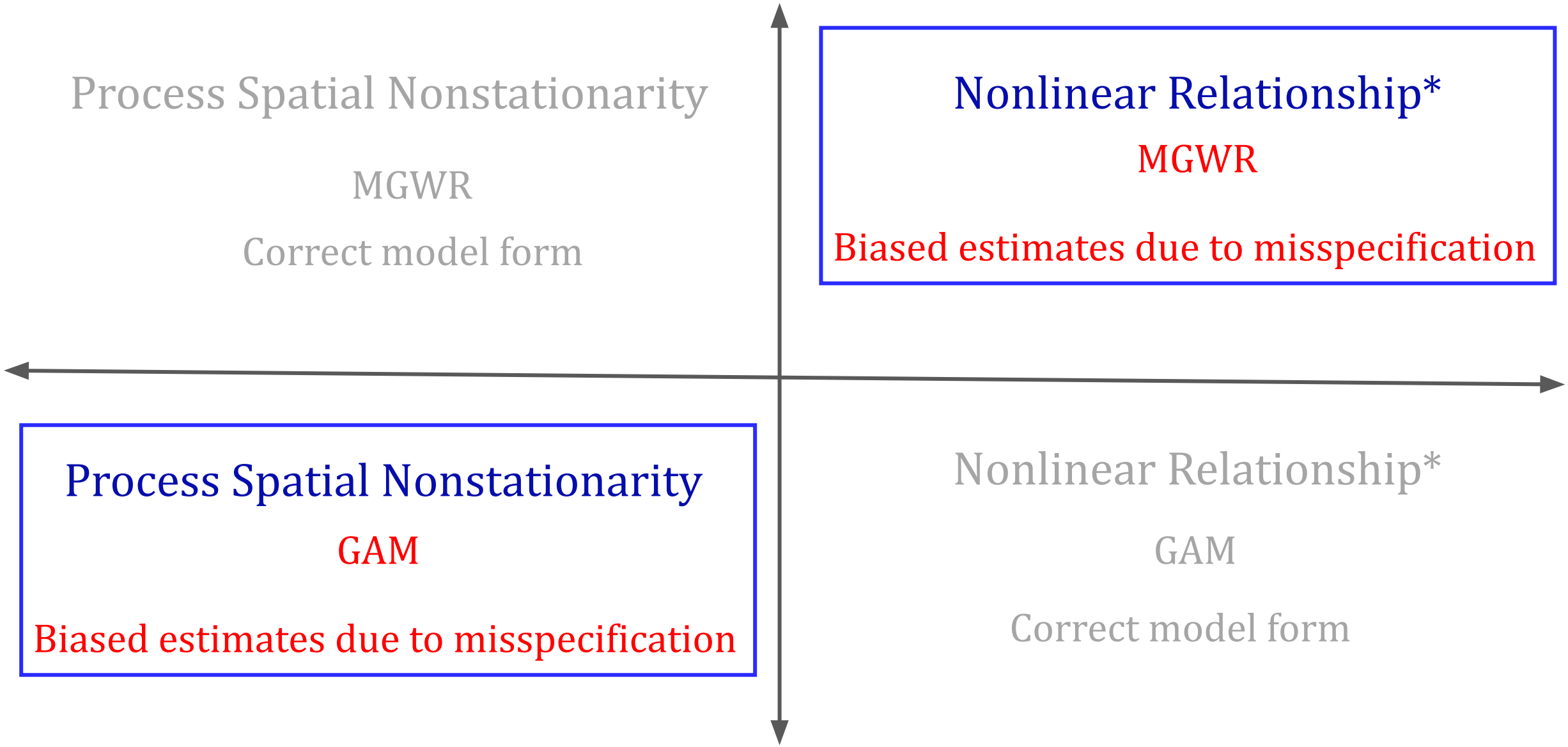
Residuals from GAM and MGWR



Moran's I value: Not significant

Process Nonstationarity Not Detected

Research describes **diagnostic tests** to detect these misspecifications



* Assuming the covariate is spatially varying

4. Intellectual merits and Broader impacts

- **Misinterpreting local estimates** due to misspecified covariate functional form plague the local analysis literature

- This **simple test must be a part of diagnostics** in local modeling

- **Expands the tools and tests within local spatial analysis** to enable stronger, more plausible interpretations

Thank you!

Any Questions?

Extra slides

Background

Process spatial non-stationarity

Variable: Square footage of Living area

1 Sq. Ft. increase in living area

Increase in house price by

~ $1000

Background

Process spatial non-stationarity

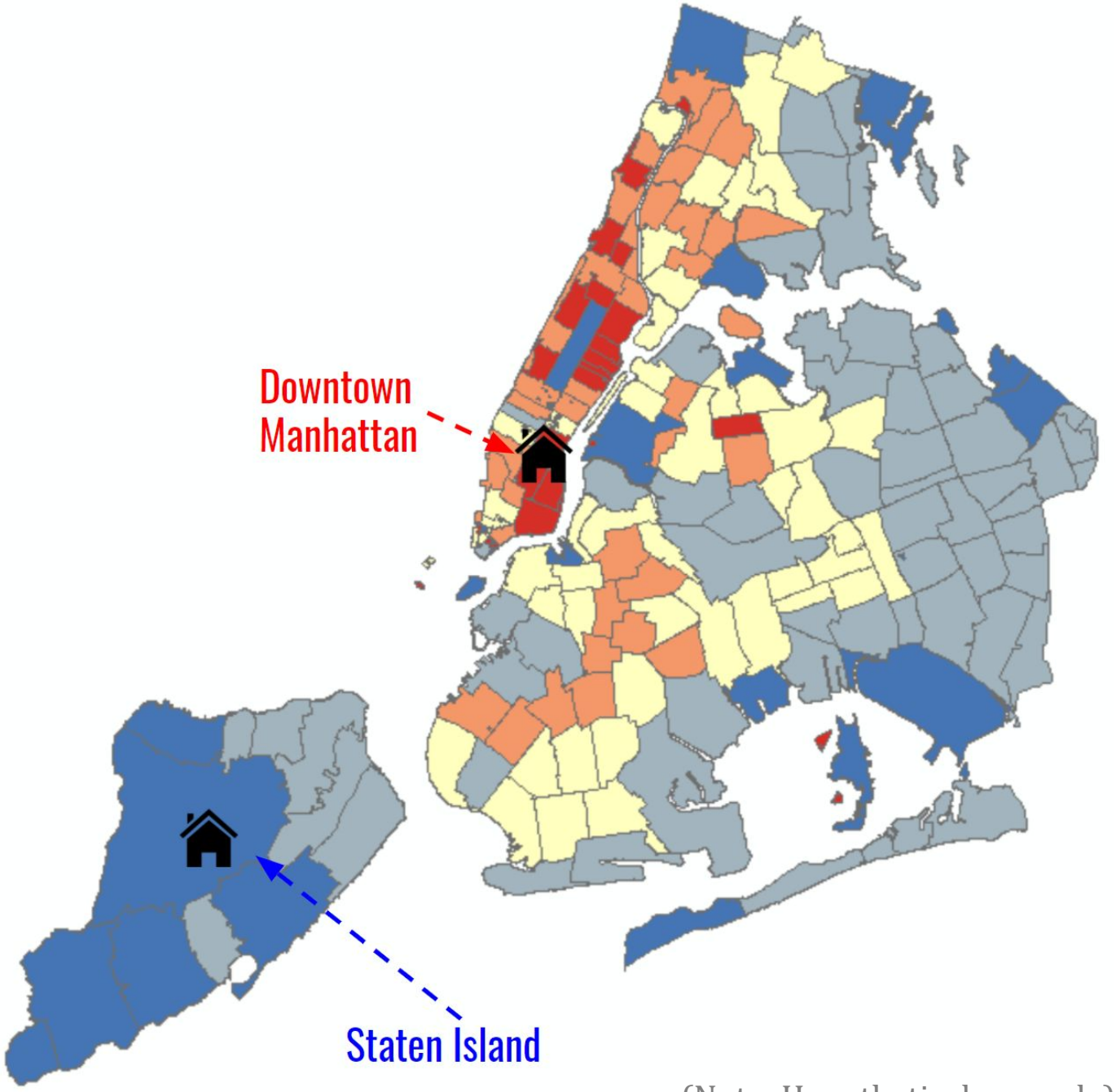
Variable: Square footage of Living area

1 Sq. Ft. increase in living area

Increase in house price by

$1600 in Downtown Manhattan

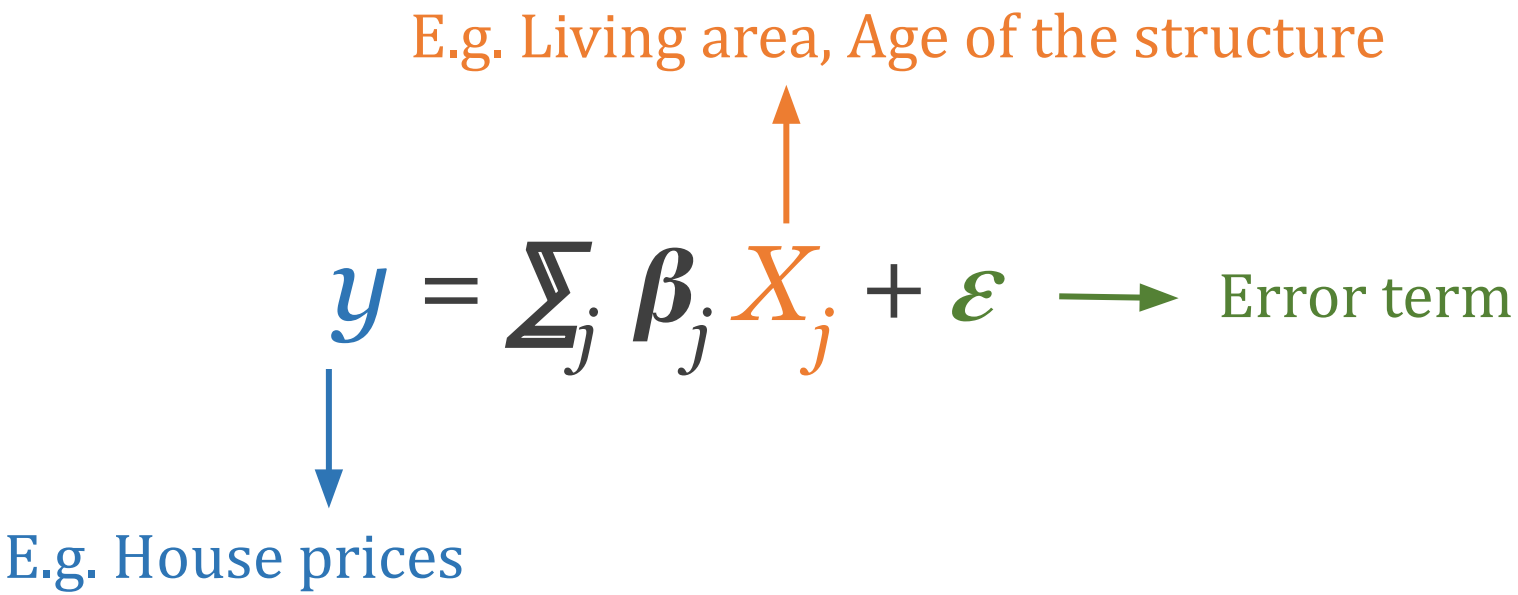
And by **$450 in Staten Island**



(Note: Hypothetical example)

Background

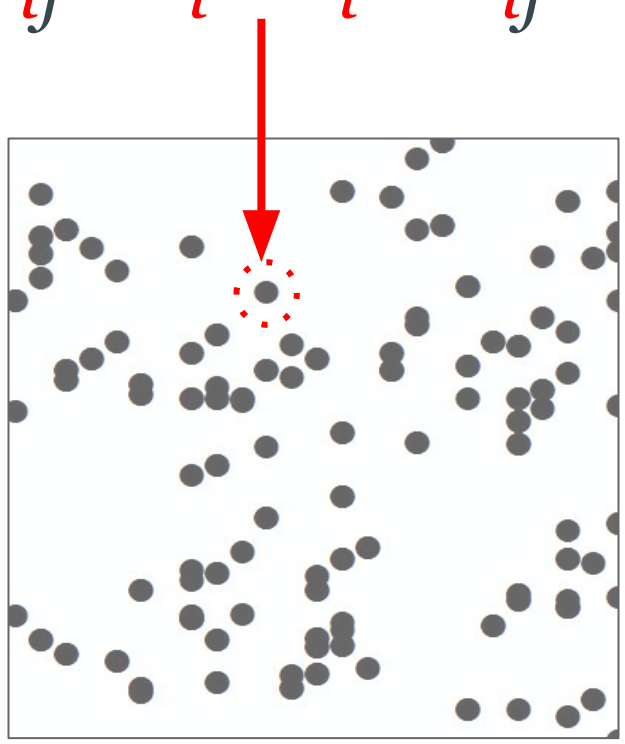
Local models within the regression framework



Background

Geographically Weighted Regression (GWR)

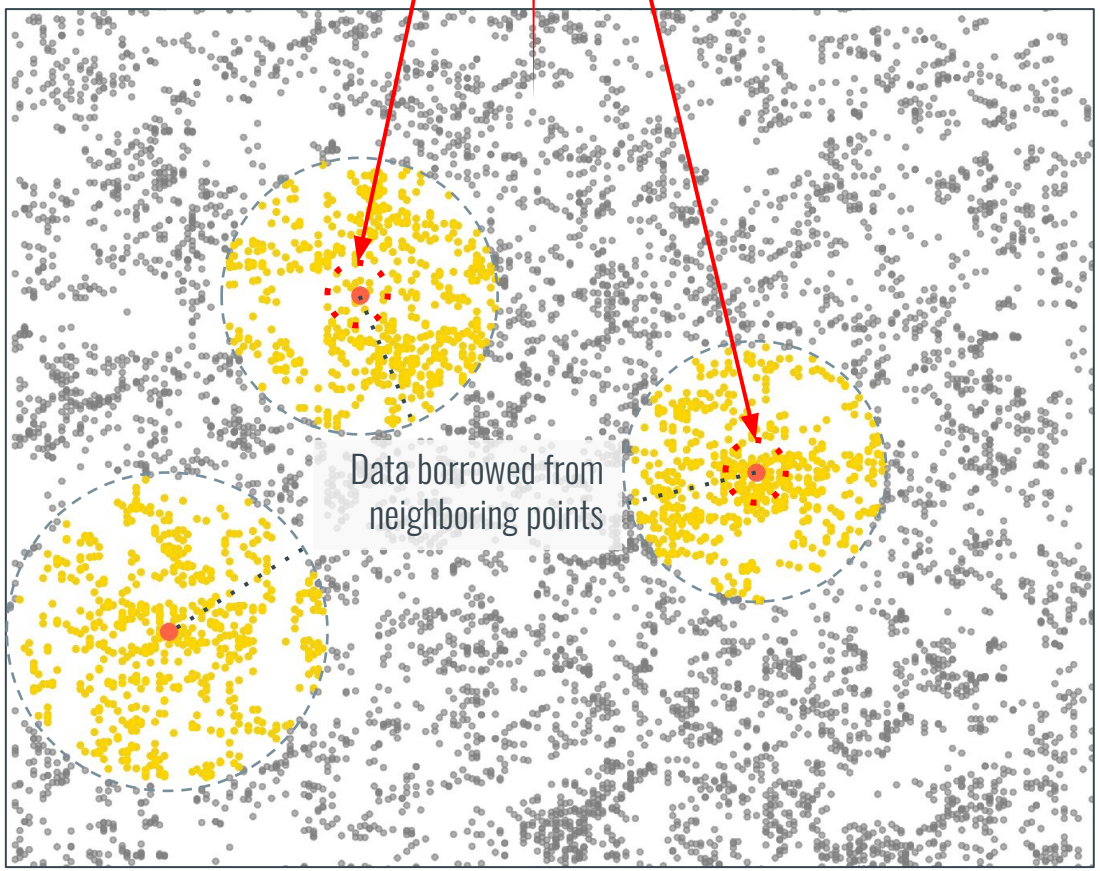
$$y_i = \sum_j \beta_{ij}(u_i, v_i) X_{ij} + \varepsilon_i$$



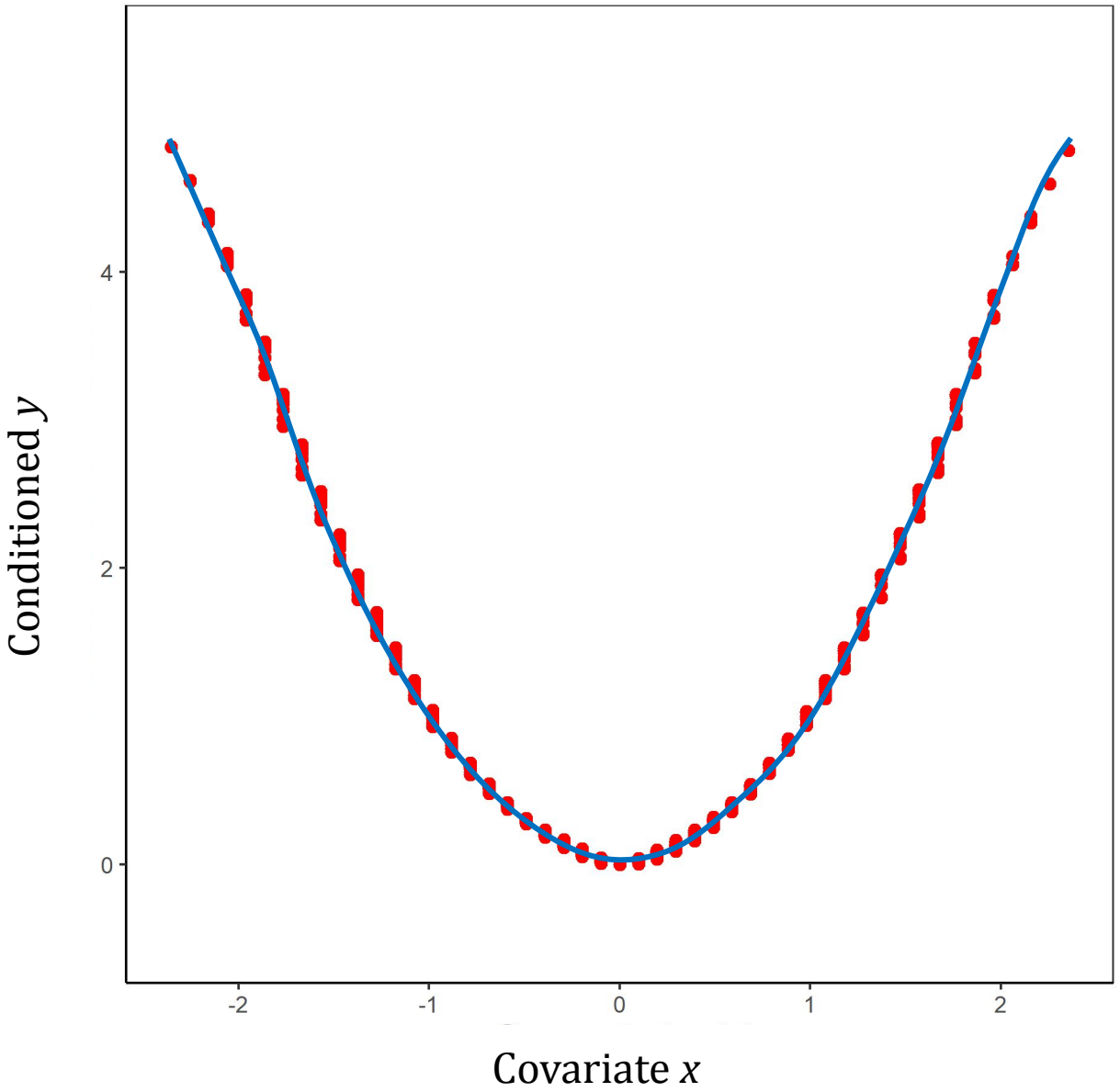
Background

Data borrowing

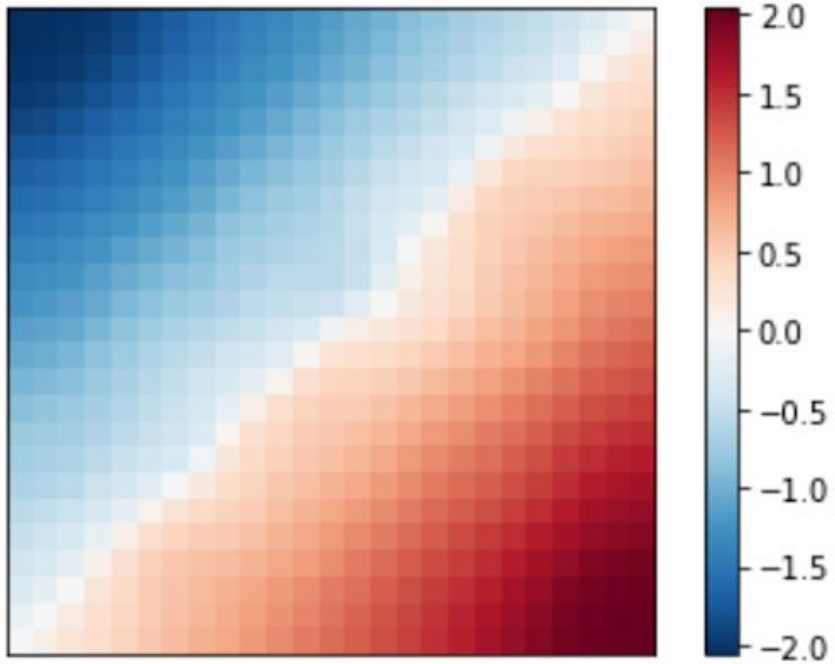
$$y_i = \sum_j \beta_{ij}(u_i, v_i) X_{ij} + \varepsilon_i$$



True conditioned relationship between x and y is:



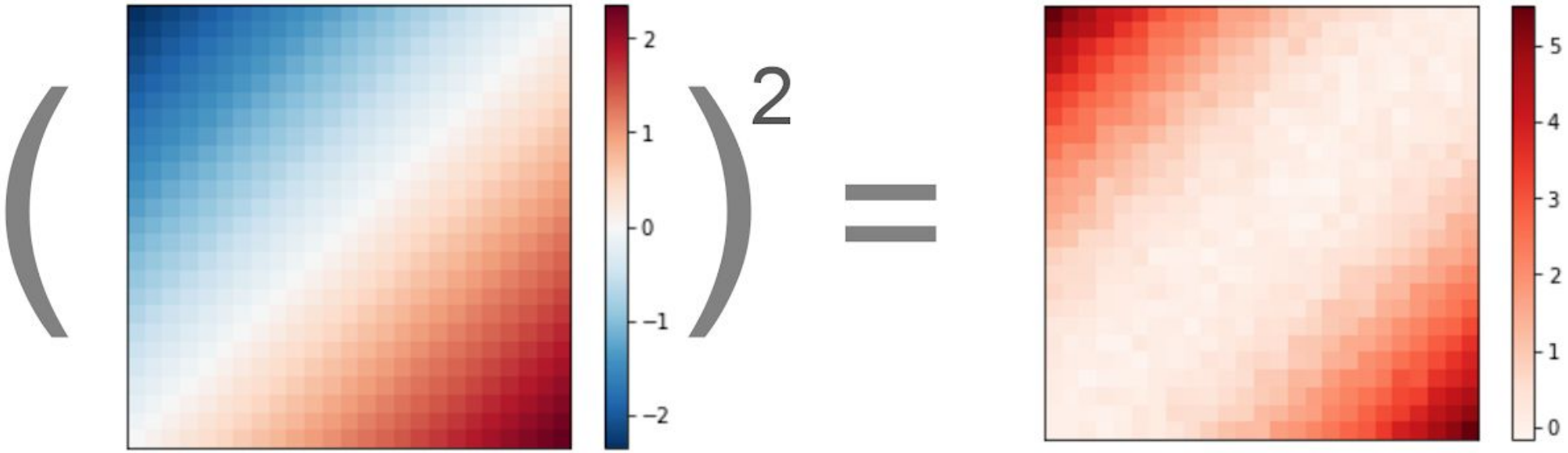
When a local model (e.g. MGWR) is used in such a situation:



Parameter estimates from MGWR

- Without examining for the possibility of nonlinearity, the spatial variation in this surface would then be incorrectly interpreted to represent process spatial non-stationarity.
- How do we differentiate such situations from presence of non-stationarity in processes?

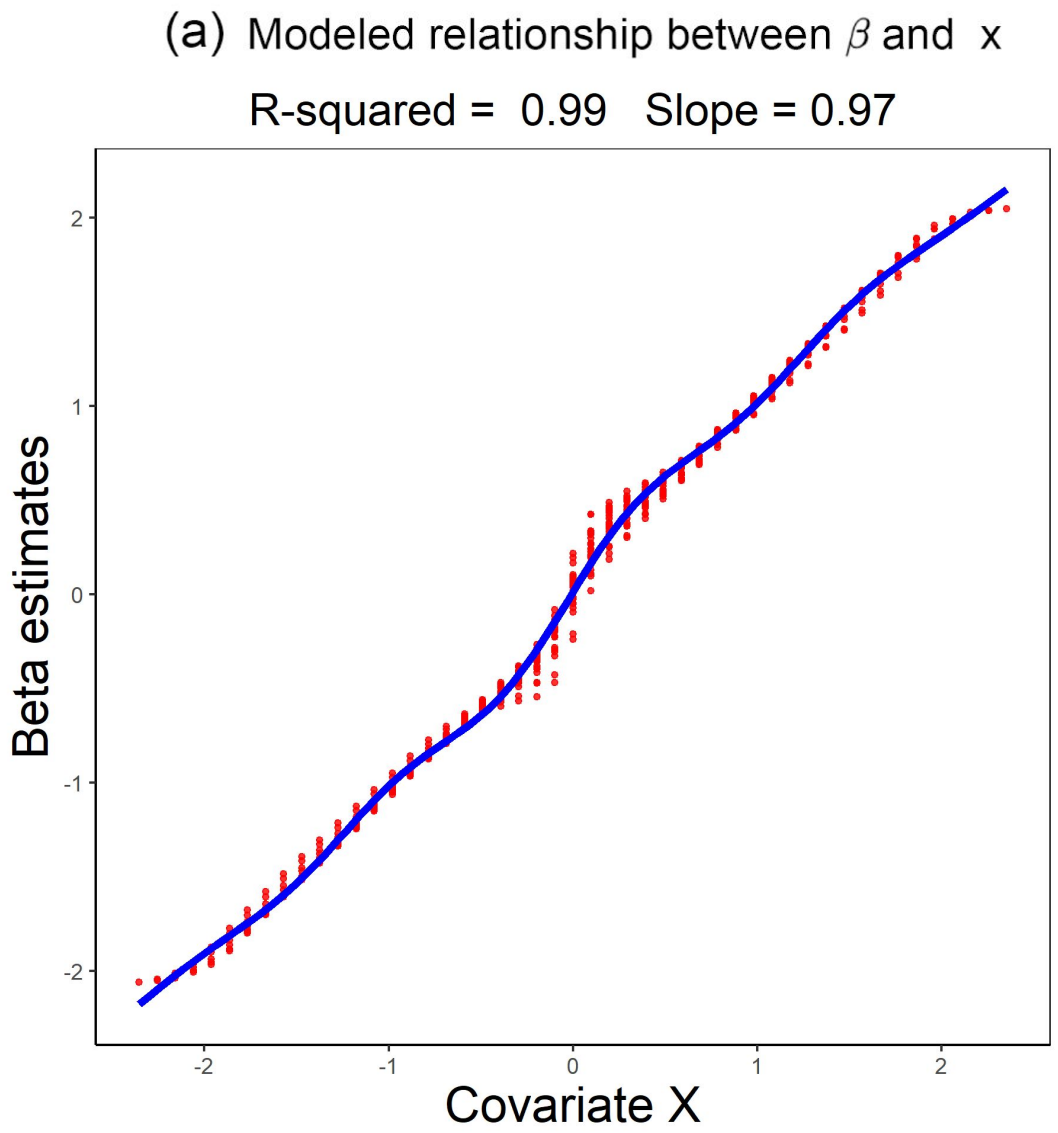
Suppose the response variable is constructed by squaring the covariate *x*

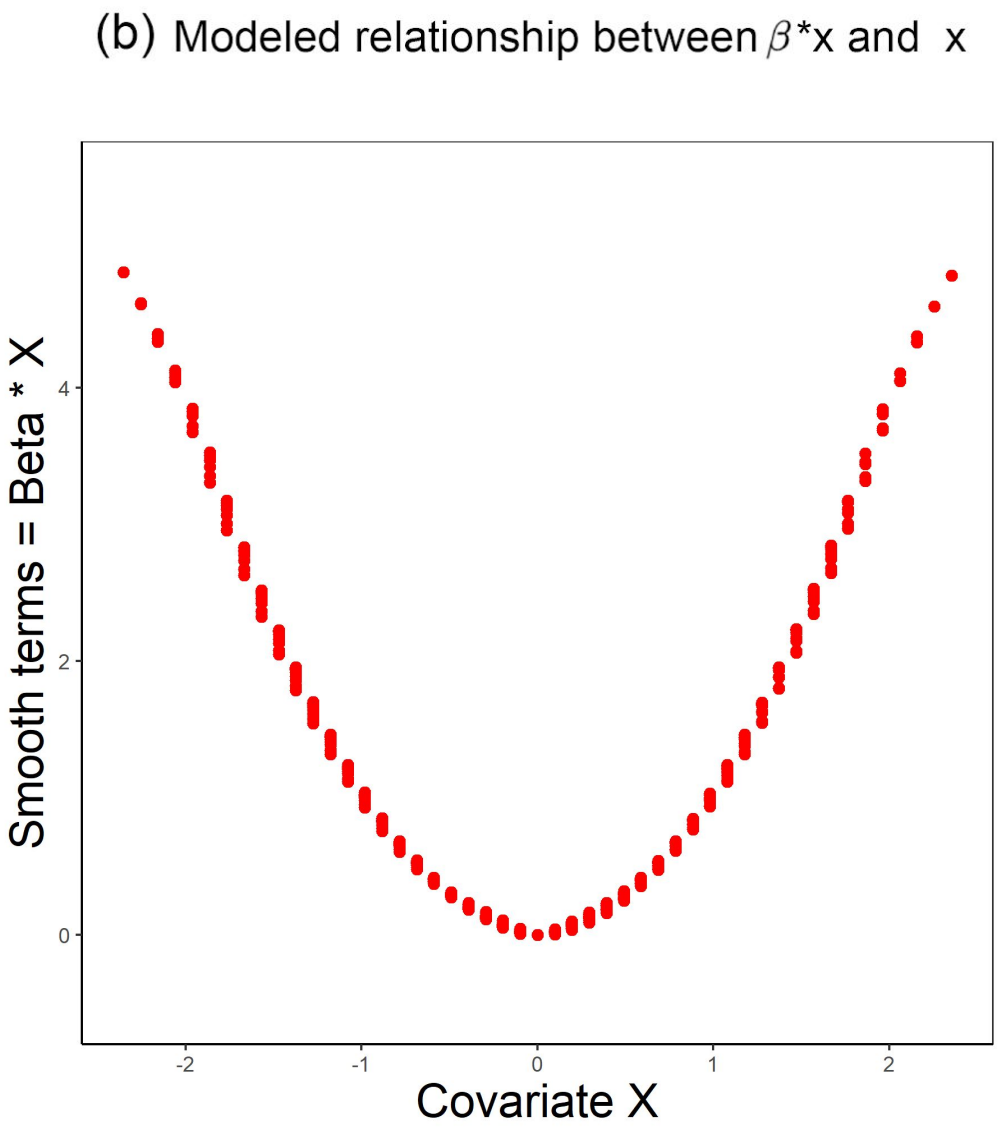


(a). Covariate x

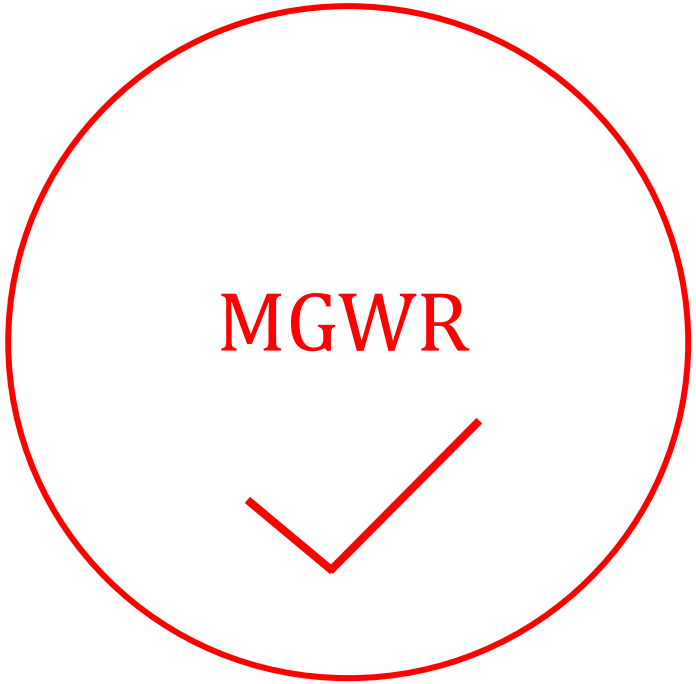
(b). Response variable y

Implementing the Diagnostic Test



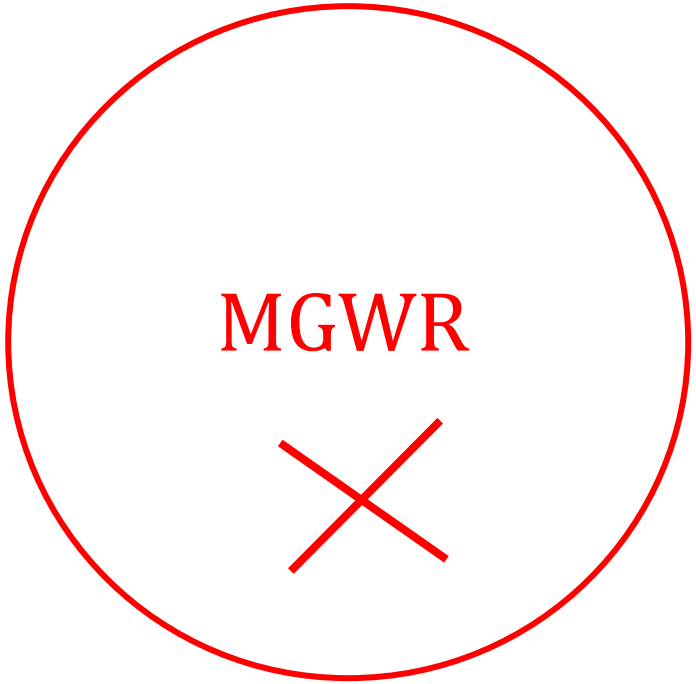


Conclusions



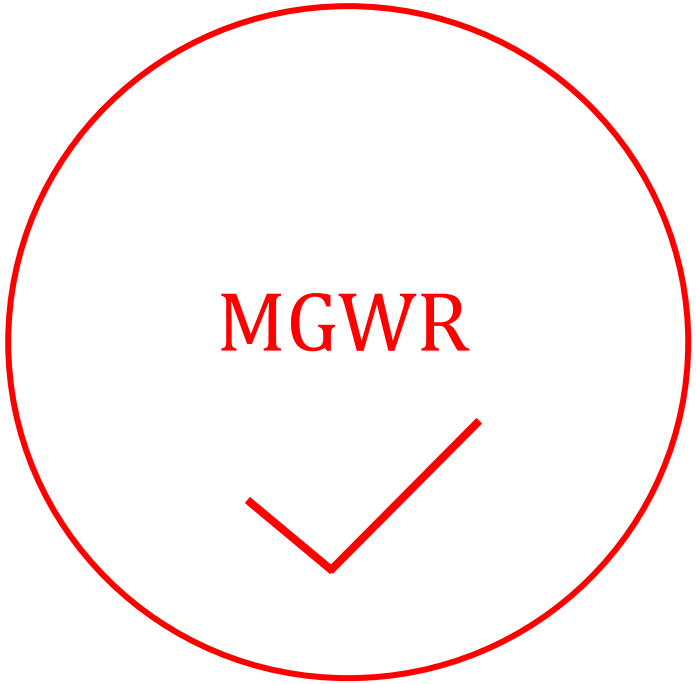
Relationships being examined are linear and spatially stationary

Conclusions



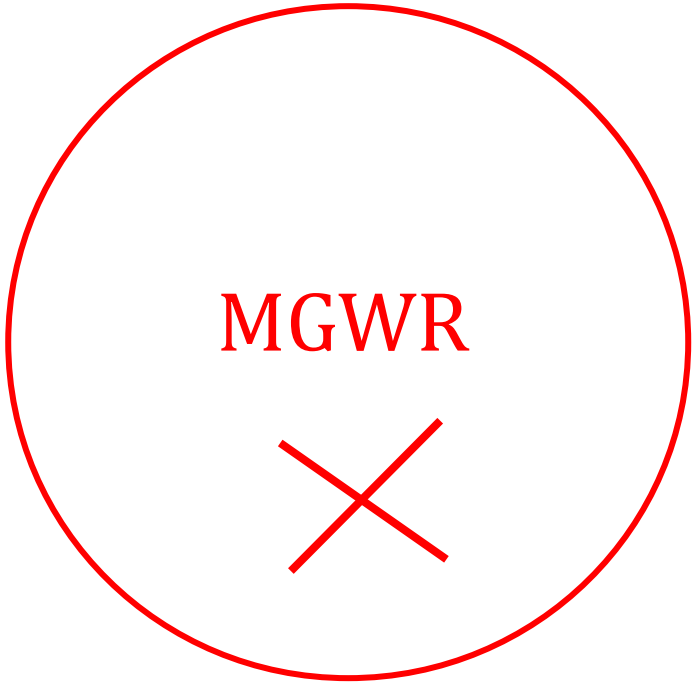
Relationships being examined are non-linear and spatially stationary

Conclusions



Relationships being examined are linear and spatially nonstationary

Conclusions



Relationships being examined are nonlinear and spatially nonstationary

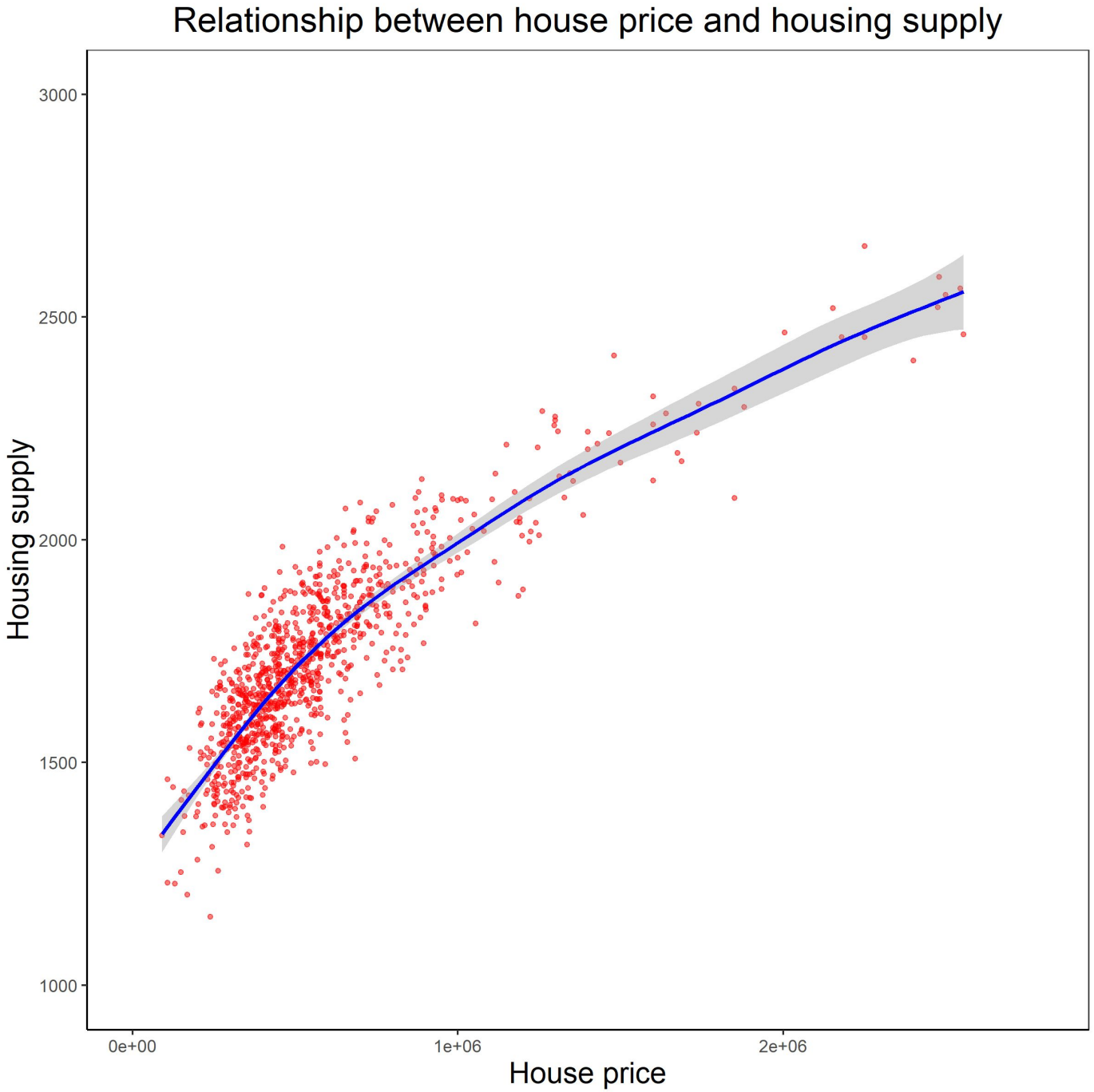
Background

Why is this problem important?

$$housing\ supply = \beta_{0(u_i,v_i)} + \beta_{1(u_i,v_i)}(house\ price)$$

Background

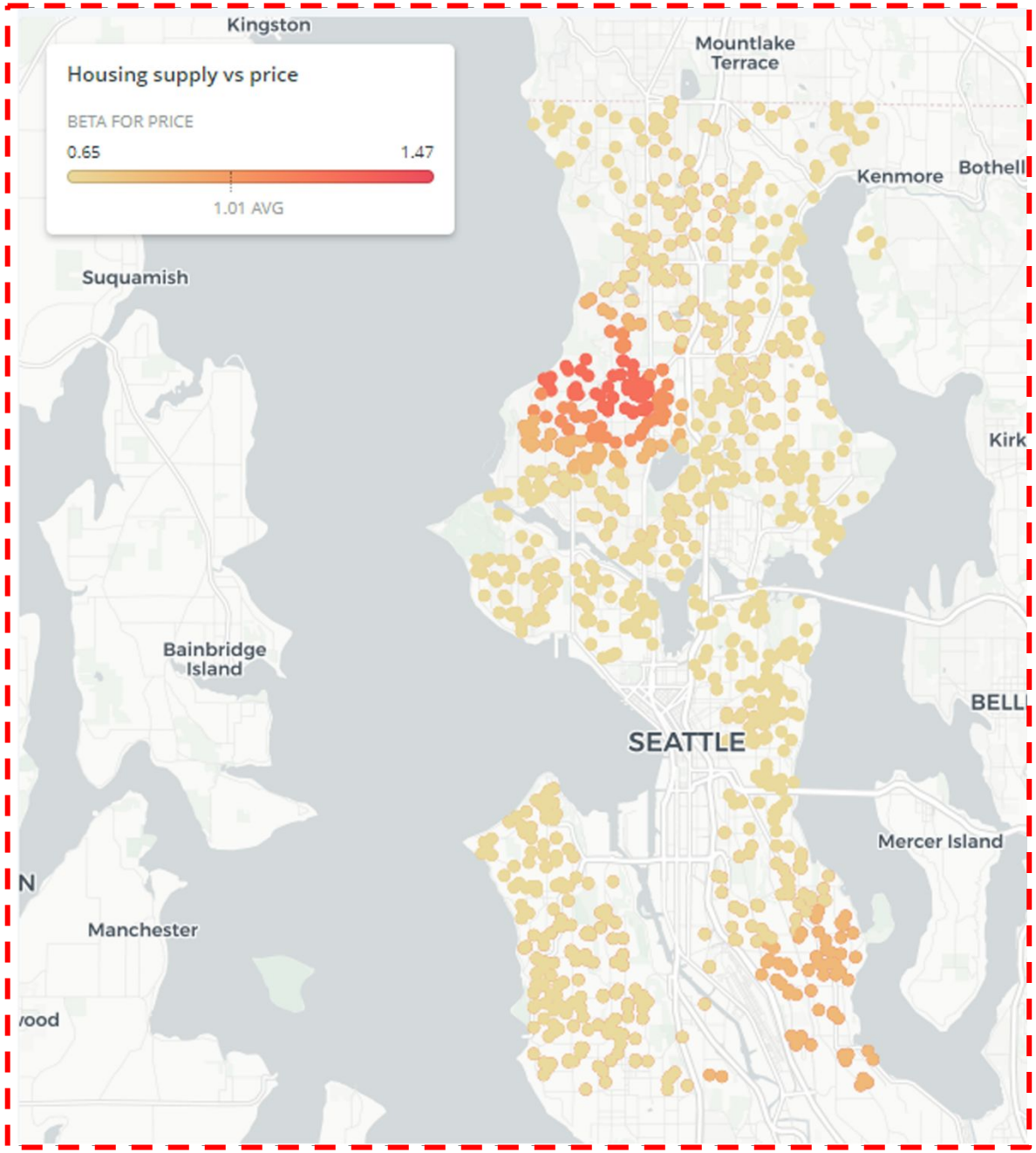
True relationship



Background

Estimated relationship

$$housing\ supply = \beta_{0(u_i,v_i)} + \beta_{1(u_i,v_i)}(\sqrt{house\ price})$$



Empirical Application

US Voting Patterns

Model of the county-level percentage of Democratic vote share [1]

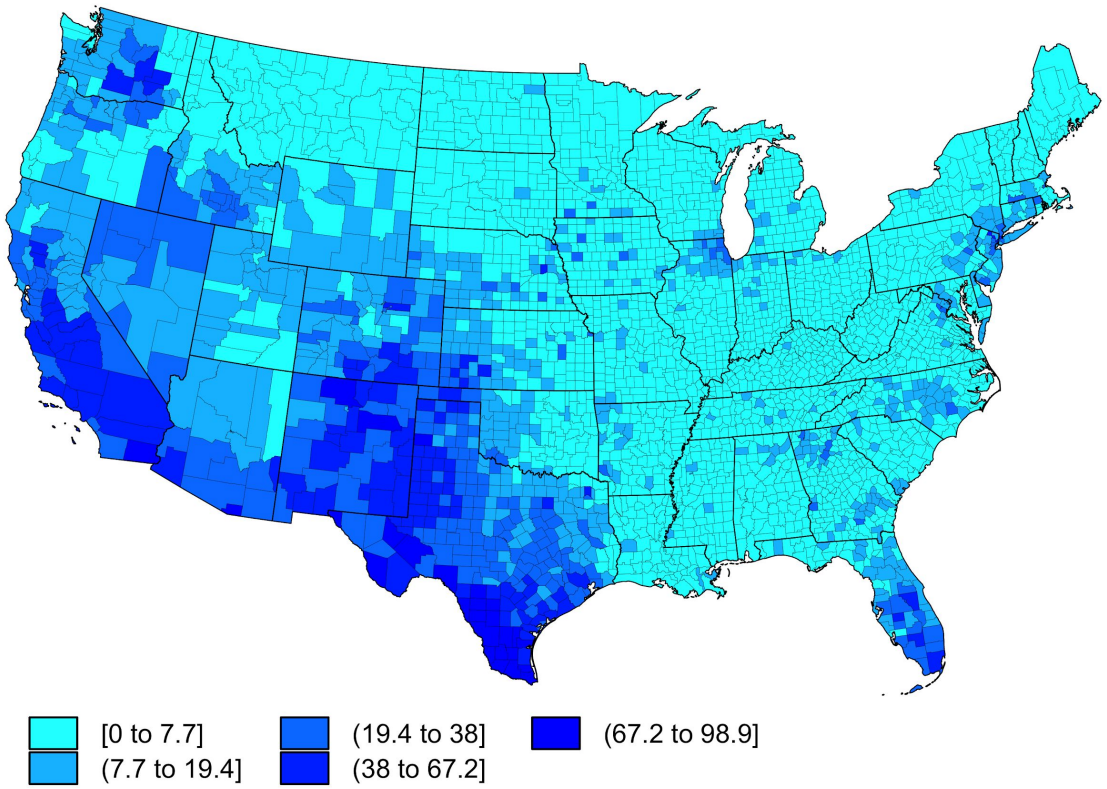
$$pct\ dem = \beta_0 + \beta_1\ sex\ ratio + \beta_2\ pct\ black + \beta_3\ pct\ hispanic + \beta_4\ pct\ bachelors + \beta_5\ median\ household\ income + \beta_6\ pct\ elderly + \beta_7\ pct\ age\ 18-29 + \beta_8\ Gini + \beta_9\ pct\ manufacturing\ employed + \beta_{10}\ ln\ pop\ density + \beta_{11}\ pct\ 3^{rd}\ party\ vote + \beta_{12}\ turnout + \beta_{13}\ pct\ foreign\ born + \beta_{14}\ pct\ insured$$

[1]Fotheringham, A. S., Li, Z., & Wolf, L. J. (2021). Scale, Context, and Heterogeneity: A Spatial Analytical Perspective on the 2016 US Presidential Election. *Annals of the American Association of Geographers*, doi: 10.1080/24694452.2020.1835459

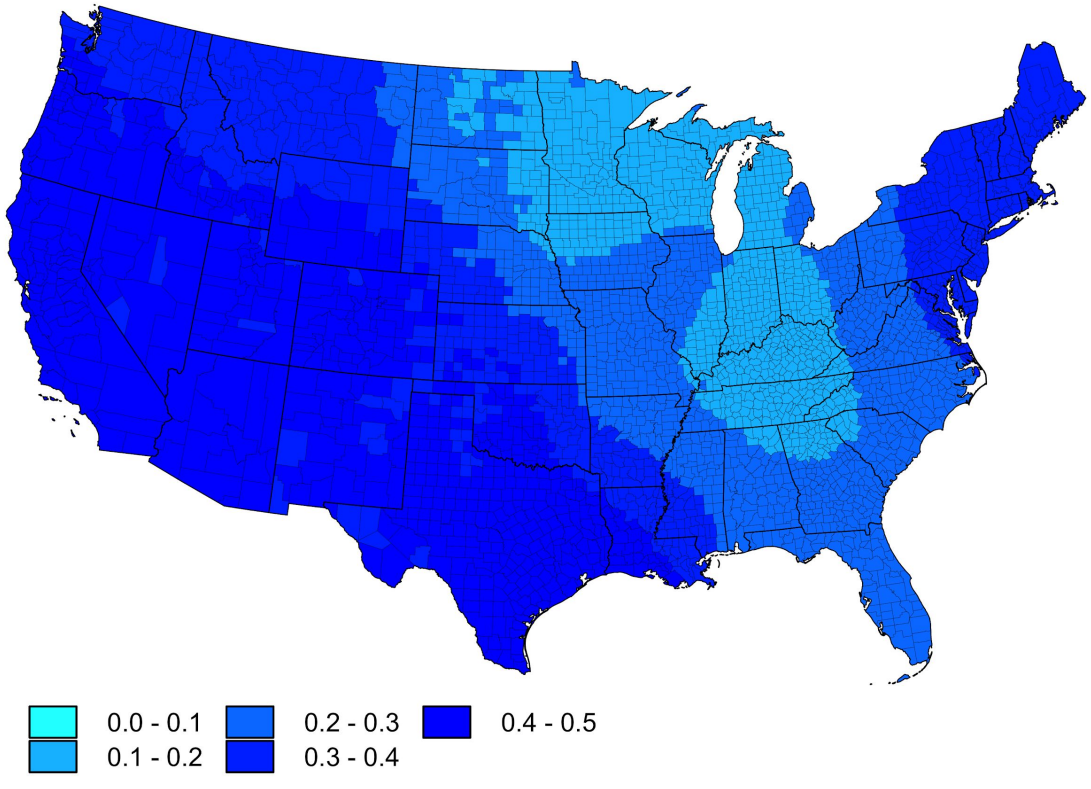
Empirical Application

US Voting Patterns

County-level Percentage of Hispanic

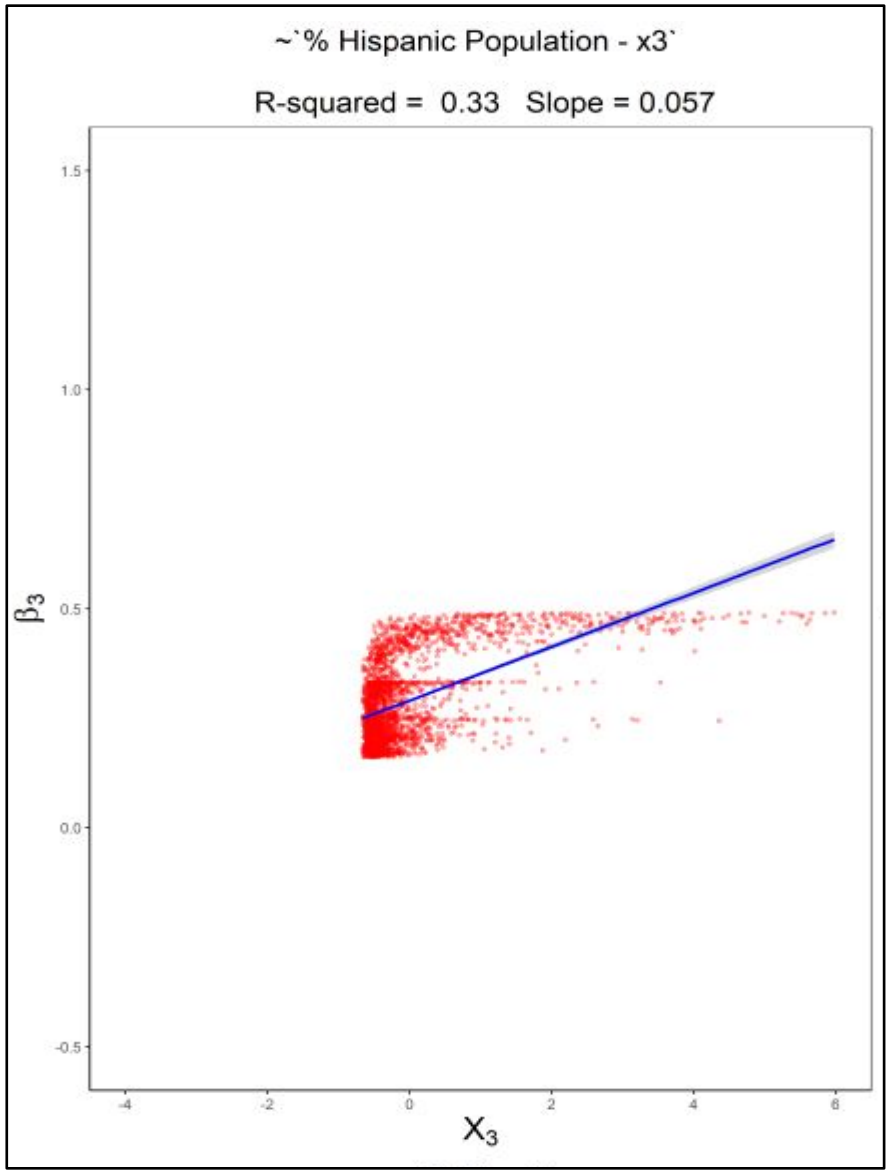


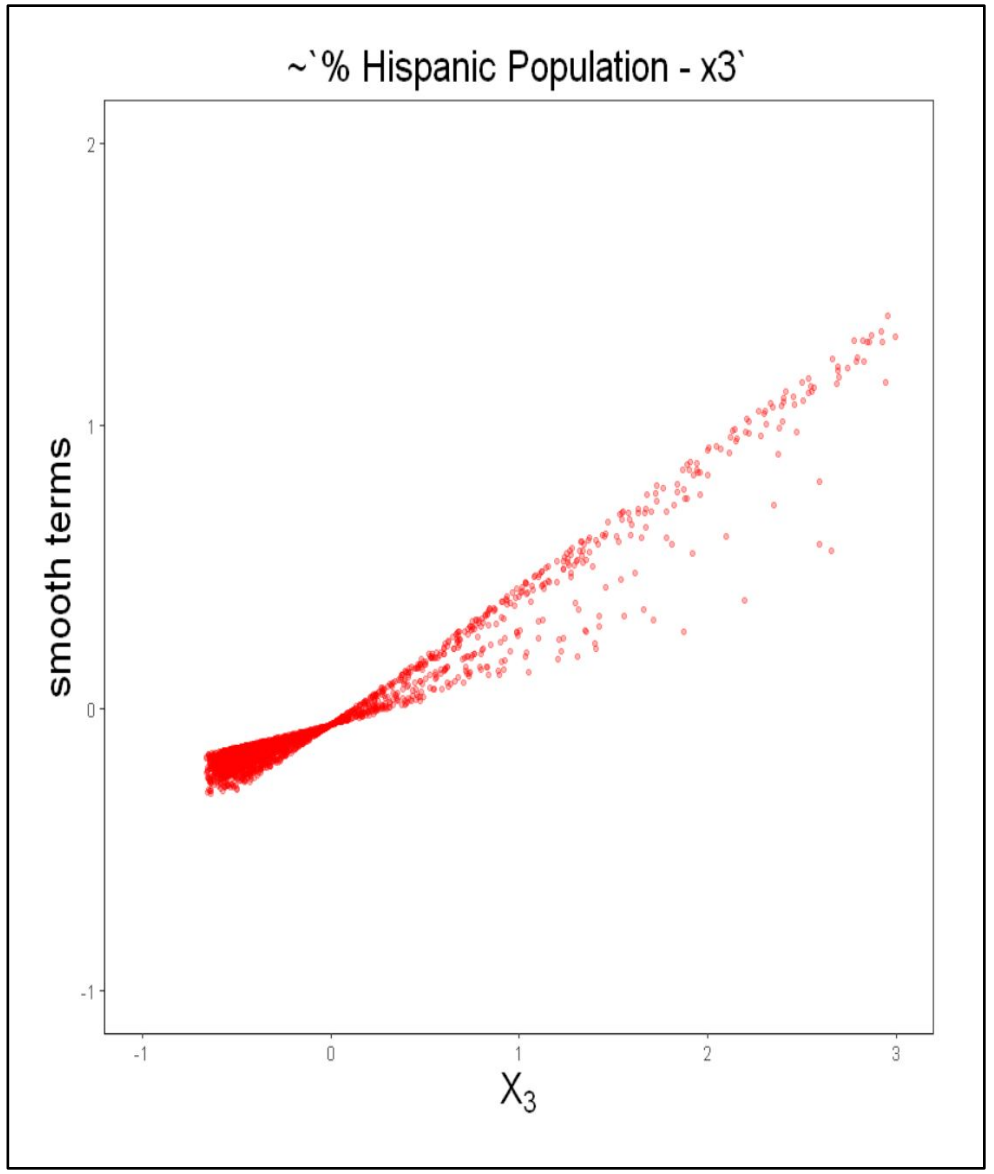
MGWR Local Parameter Estimates for Percentage of Hispanic



Empirical Application

US Voting Patterns





Empirical Application

US Voting Patterns

Using % Hispanic population^2 as a predictor