# Can Conventional Measures Identify Geographically Varying Mixed Regression Relationships? A Simulation-based Analysis of Locally Weighted Regression

Aaron Swoboda

Carleton College

aswoboda@carleton.edu

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \epsilon$$

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \epsilon$$

Locally Weighted Regression (LWR)
to the Rescue?

# OLS

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

# LWR

$$\hat{\beta}(location_i) = (X'W(location_i)X)^{-1}(X'W(location_i)Y)$$

# OLS

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

# LWR

$$\hat{\beta}(location_i) = (X'W(location_i)X)^{-1}(X'W(location_i)Y)$$

$$w_{jj} = \left[1 - \left(\frac{d_{ij}}{d_{ik}}\right)^2\right]^2 \text{ if } d_{ij} < d_{ik}, \text{ otherwise} = 0,$$

Bandwidths are commonly selected with…

Leave One Out Cross Validation

Akaike Information Criterion

Generalized Cross Validation

Standardized Cross Validation

$$LOOCV = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(y - \hat{y}_{\neq i})^2},$$

$$SCV_i(k) = \frac{(y_i - \hat{y}_{\neq i}(k))^2}{\sum_k (y_i - \hat{y}_{\neq i})^2}$$

$$SCV(k) = \sum_i SCV_i(k)$$

S Farber and A Páez. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, 9(4):371–396, 2007.

$$GCV = n * \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(n - v_1)^2},$$

$$v_1 = \text{“effective number of parameters”}$$

$$AIC = 2 * n * ln(\hat{\sigma}) + n * ln(2 * \pi) + n * \frac{n + v_1}{n - 2 - v_1}$$

“Global” OLS GGG

“Local”Regression LLL

| | |
|---|---|
| "Global" OLS | GGG |
| | LGG |
| | GLG |
| | GGL |
| Mixed Models | LLG |
| | LGL |
| | GLL |
| "Local" Regression | LLL |

# The Researcher's Problem

Choose a bandwidth <u>and</u> a model

# The Researcher's Problem

Choose a bandwidth <u>and</u> a model

Can we do both with conventional metrics?

# Experiment Data Generation Process

$$Y_i = \beta_0(East_i, North_i) + \beta_1(East_i, North_i) * X_{1i} + \beta_2(East_i, North_i) * X_{2i} + \epsilon_i$$

$$n \in \{50, 100, 200, 400, 800\}$$

$$\sigma^2 \in \{0.25, .5, 1, 2, 3\}$$

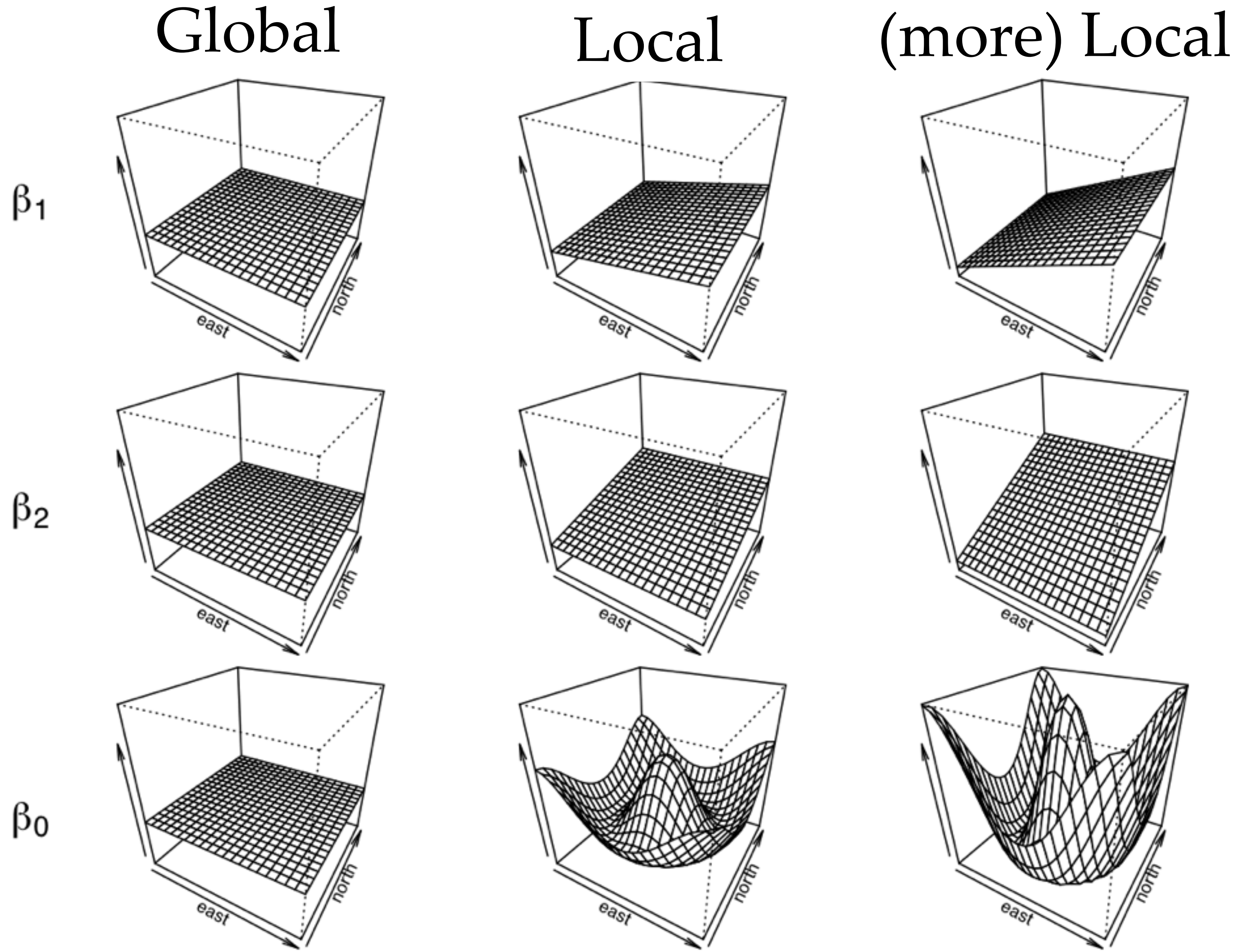
$$X_1 \sim u[0,1]$$

$$X_2 \sim u[0,1]$$

$$East \sim u[0,1]$$

$$North \sim u[0,1]$$

# Coefficient Spatial Variation



# With Our Data…

$$[Y, X_1, X_2, East, North]$$

Estimate all models

| |
|---|
| GGG |
| LGG |
| GLG |
| GGL |
| LLG |
| LGL |
| GLL |
| LLL |

7 bandwidths each

50 combinations total

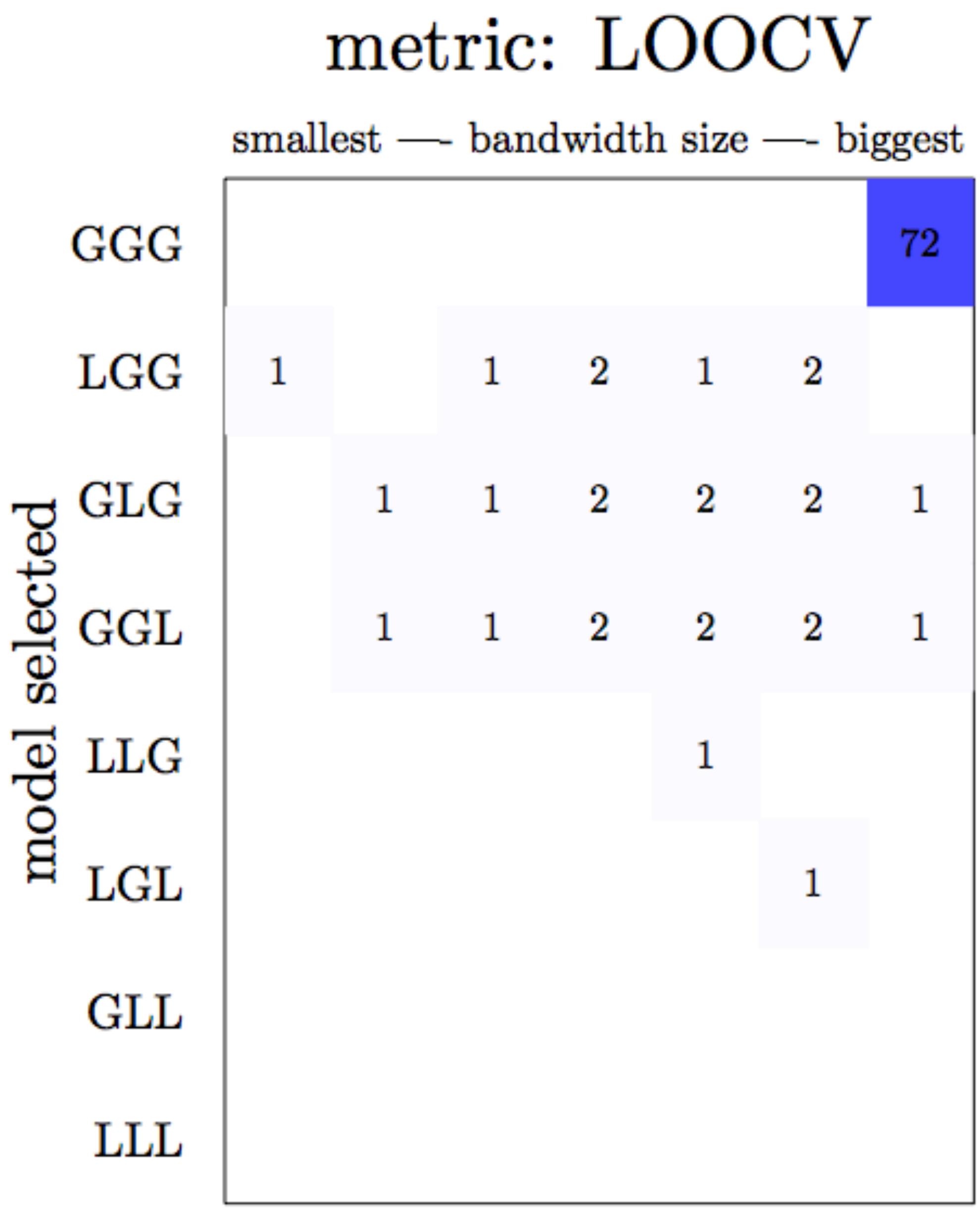
Calculate the values of the four metrics (LOOCV, GCV, SCV, AIC)

Is the model with the optimized metric value the correct model?

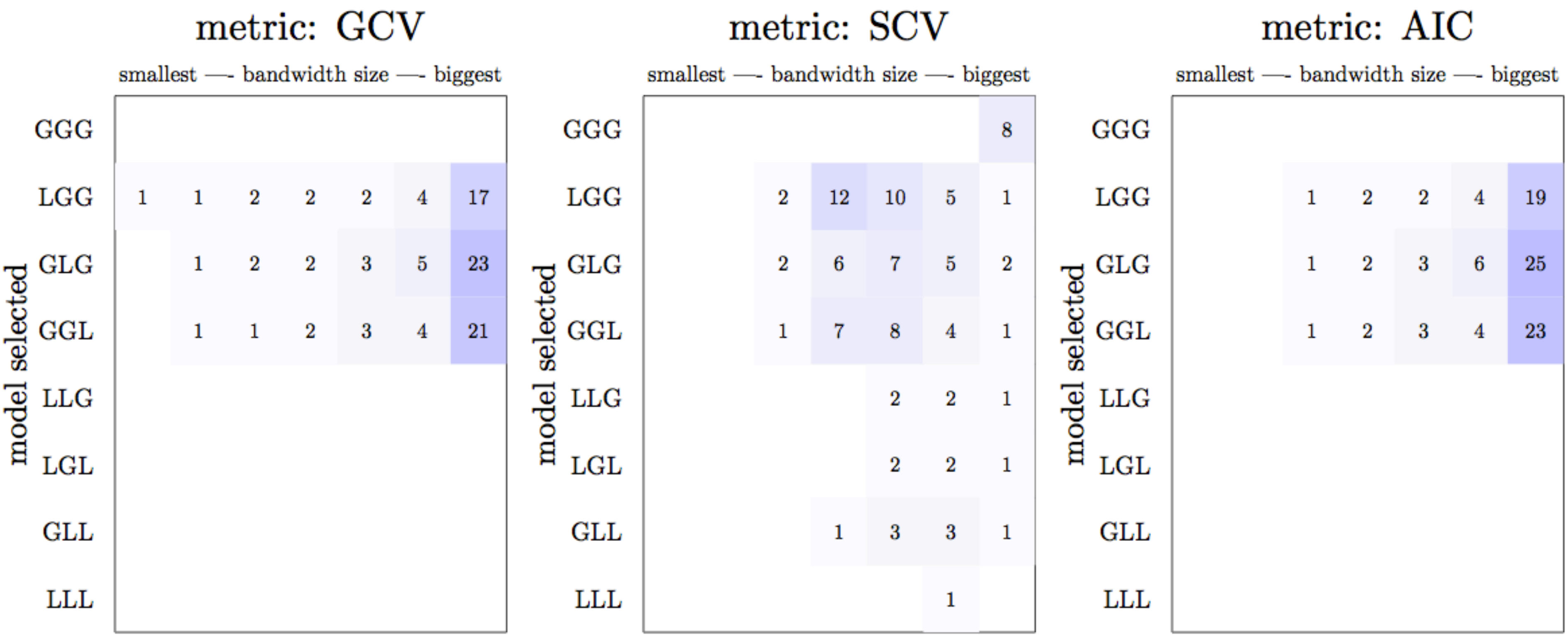
# Start Simple… True Model: GGG

| Model Selected | Metric: LOOCV | GCV | SCV | AIC | |
|---|---|---|---|---|---|
| GGG | 72 | 0 | 8 | 0 | 3/3 Correct |
| LGG | 7 | 28 | 29 | 28 | |
| GLG | 8 | 36 | 22 | 37 | 2/3 Correct |
| GGL | 8 | 33 | 22 | 34 | |
| LLG | 1 | 1 | 5 | 0 | |
| LGL | 2 | 1 | 5 | 1 | 1/3 Correct |
| GLL | 1 | 1 | 8 | 0 | |
| LLL | 0 | 0 | 1 | 0 | 0/3 Correct |
| | 100 | 100 | 100 | 100 | |

# True Model: GGG



# True Model: GGG



# True Model: GGG

| Model Selected | Coefficient RMSE $\widehat{\beta_0}$ | $\widehat{\beta_1}$ | $\widehat{\beta_2}$ |
|---|---|---|---|
| GGG | 6 | 6 | 7 |
| LGG | 3 | 23 | 24 |
| GLG | 22 | 4 | 24 |
| GGL | 23 | 23 | 3 |
| LLG | 5 | 3 | 33 |
| LGL | 5 | 36 | 2 |
| GLL | 32 | 4 | 4 |
| LLL | 4 | 2 | 3 |
| | 100 | 100 | 100 |

# True Model: GGG

### metric: B0RMSE

smallest —- bandwidth size —- biggest

| model selected | | | | | | | |
|---|---|---|---|---|---|---|---|
| GGG | | | | | | | 6 |
| LGG | | | | | | 1 | 3 |
| GLG | 9 | 3 | 2 | 2 | 2 | 2 | 1 |
| GGL | 9 | 3 | 3 | 3 | 2 | 1 | 1 |
| LLG | | | | | | | 4 |
| LGL | | | | | | | 4 |
| GLL | 14 | 5 | 4 | 4 | 3 | 2 | 1 |
| LLL | | | | | | | 4 |

### metric: B1RMSE

smallest —- bandwidth size —- biggest

| model selected | | | | | | | |
|---|---|---|---|---|---|---|---|
| GGG | | | | | | | 6 |
| LGG | 8 | 4 | 3 | 3 | 2 | 2 | 1 |
| GLG | | | | | | | 3 |
| GGL | 10 | 3 | 3 | 2 | 2 | 2 | 1 |
| LLG | | | | | | | 3 |
| LGL | 15 | 5 | 4 | 4 | 3 | 2 | 3 |
| GLL | | | | | | | 4 |
| LLL | | | | | | | 2 |

### metric: B2RMSE

smallest —- bandwidth size —- biggest

| model selected | | | | | | | |
|---|---|---|---|---|---|---|---|
| GGG | | | | | | | 7 |
| LGG | 9 | 4 | 3 | 2 | 2 | 2 | 2 |
| GLG | 10 | 4 | 2 | 1 | 2 | 2 | 2 |
| GGL | | | | | | | 3 |
| LLG | 12 | 5 | 4 | 4 | 3 | 3 | 2 |
| LGL | | | | | | | 2 |
| GLL | | | | | | | 4 |
| LLL | | | | | | | 3 |

# True Model: GGG

### metric: B0RMSE

smallest —- bandwidth size —- biggest

| model selected | | | | | | | |
|---|---|---|---|---|---|---|---|
| GGG | | | | | | | 6 |
| LGG | | | | | | 1 | 3 |
| GLG | 9 | 3 | 2 | 2 | 2 | 2 | 1 |
| GGL | 9 | 3 | 3 | 3 | 2 | 1 | 1 |
| LLG | | | | | | | 4 |
| LGL | | | | | | | 4 |
| GLL | 14 | 5 | 4 | 4 | 3 | 2 | 1 |
| LLL | | | | | | | 4 |

### metric: B1RMSE

smallest —- bandwidth size —- biggest

| model selected | | | | | | | |
|---|---|---|---|---|---|---|---|
| GGG | | | | | | | 6 |
| LGG | 8 | 4 | 3 | 3 | 2 | 2 | 1 |
| GLG | | | | | | | 3 |
| GGL | 10 | 3 | 3 | 2 | 2 | 2 | 1 |
| LLG | | | | | | | 3 |
| LGL | 15 | 5 | 4 | 4 | 3 | 2 | 3 |
| GLL | | | | | | | 4 |
| LLL | | | | | | | 2 |

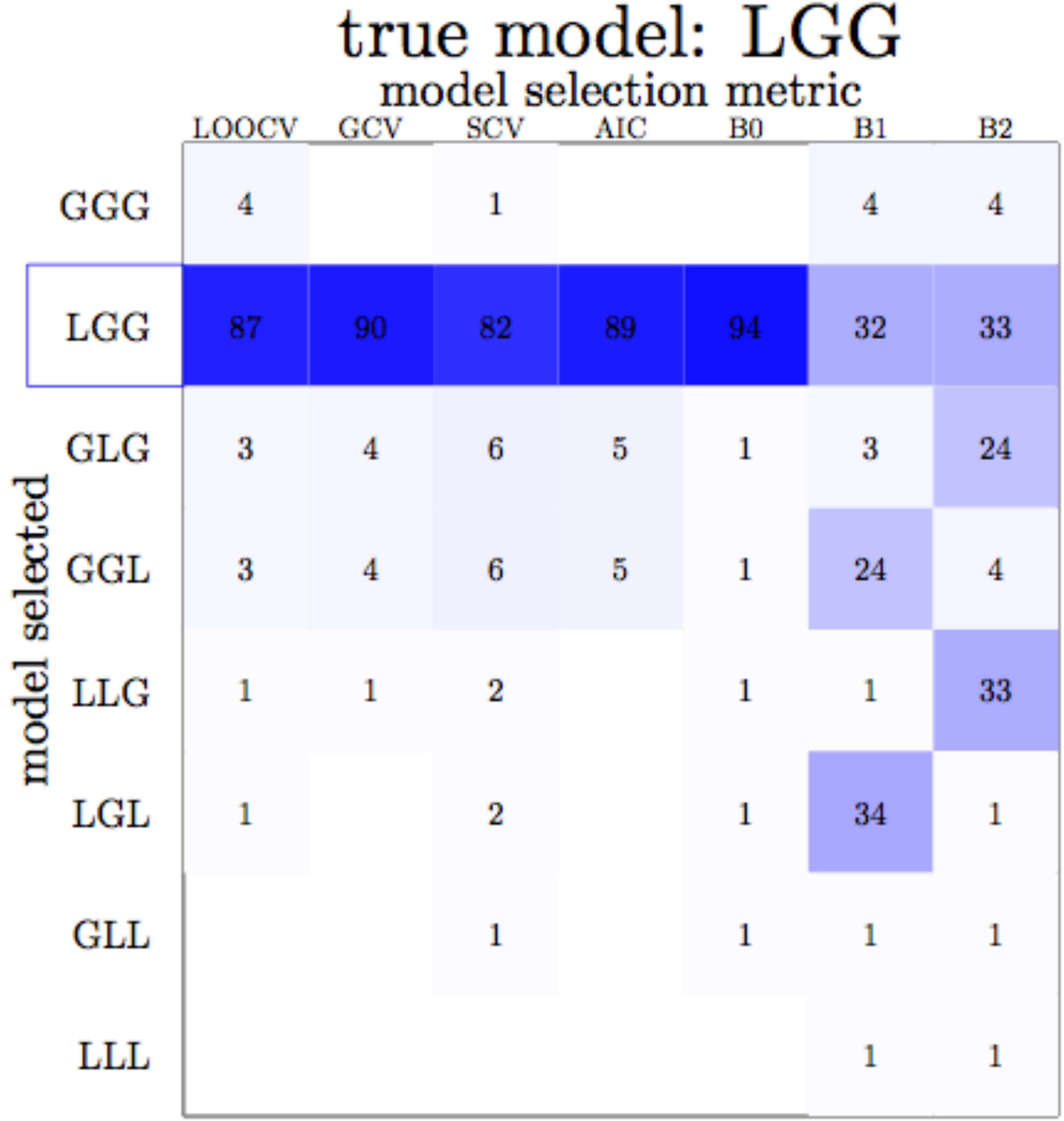
### metric: B2RMSE

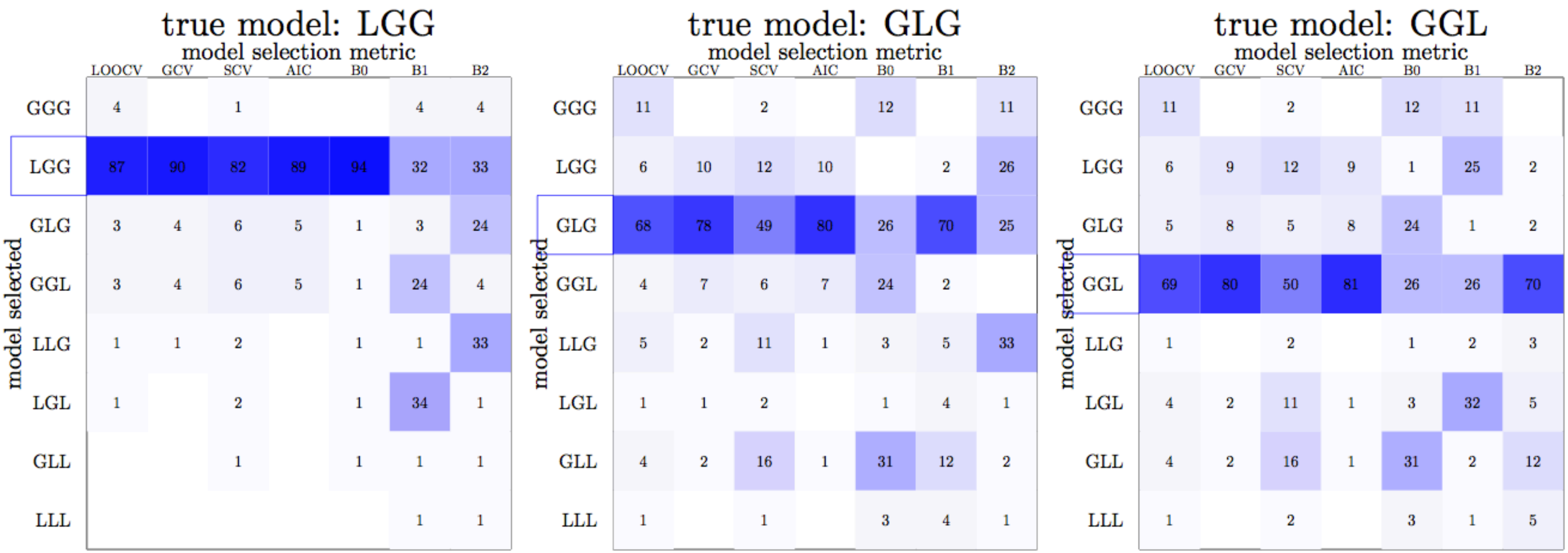
smallest —- bandwidth size —- biggest

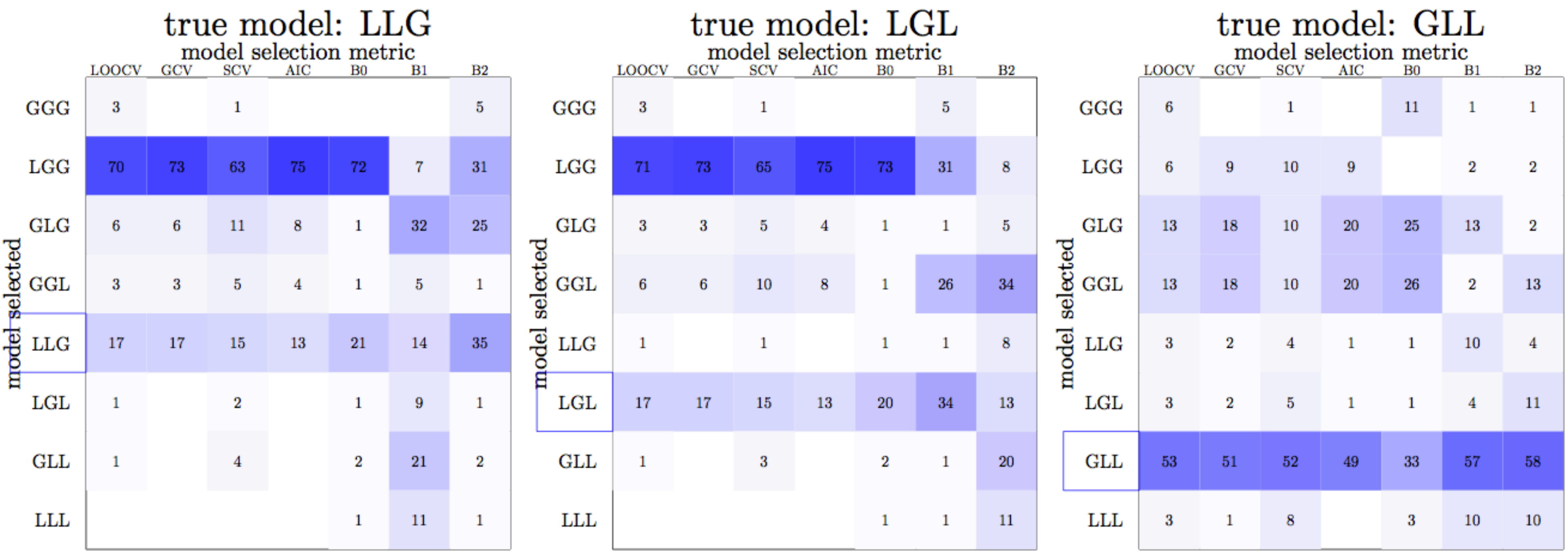
| model selected | | | | | | | |
|---|---|---|---|---|---|---|---|
| GGG | | | | | | | 7 |
| LGG | 9 | 4 | 3 | 2 | 2 | 2 | 2 |
| GLG | 10 | 4 | 2 | 1 | 2 | 2 | 2 |
| GGL | | | | | | | 3 |
| LLG | 12 | 5 | 4 | 4 | 3 | 3 | 2 |
| LGL | | | | | | | 2 |
| GLL | | | | | | | 4 |
| LLL | | | | | | | 3 |

Most accurate coefficient estimates:

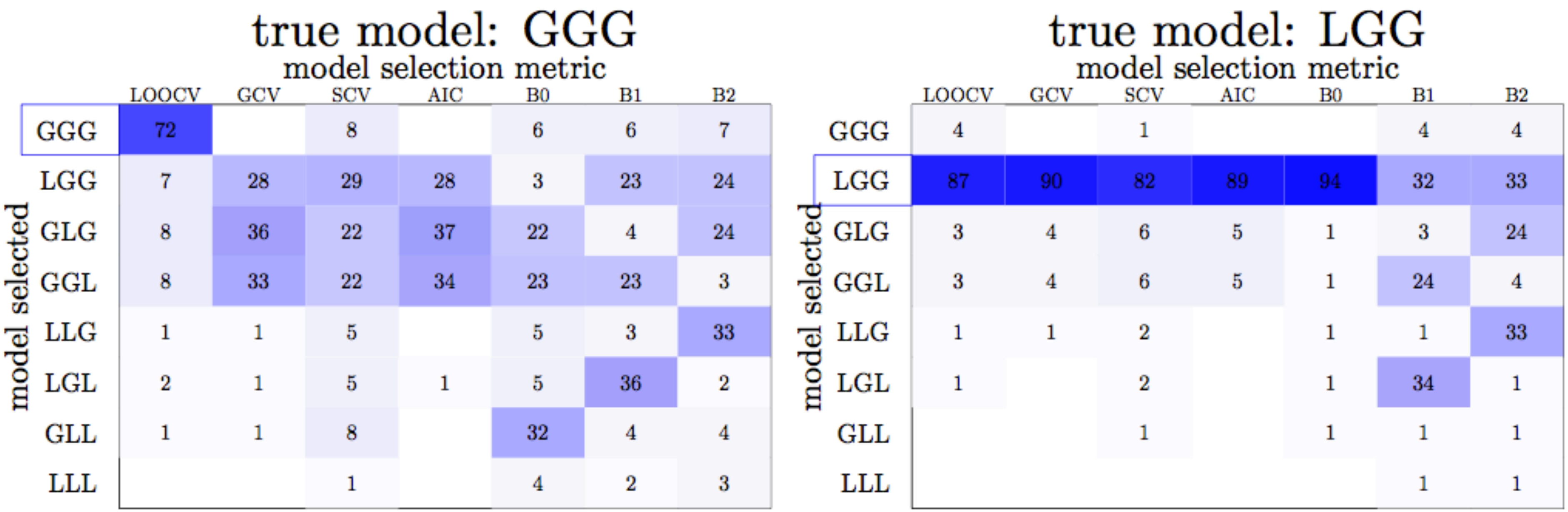
- Tend not to be the correct model
- Allow other coefficients to vary


true model: LGG
model selection metric
LOOCV
GCV
SCV
AIC
B0
B1
B2
model selected
GGG
LGG
GLG
GGL
LLG
LGL
GLL
LLL
4
1
4
4
87
90
82
89
94
32
33
3
4
6
5
1
3
24
3
4
6
5
1
24
4
1
1
2
1
1
33
1
2
1
34
1
1
1
1
1
1
1


true model: LGG
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 4 1 4 4
LGG 87 90 82 89 94 32 33
GLG 3 4 6 5 1 3 24
GGL 3 4 6 5 1 24 4
LLG 1 1 2 1 1 33
LGL 1 2 1 34 1
GLL 1 1 1 1
LLL 1 1
true model: GLG
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 11 2 12 11
LGG 6 10 12 10 2 26
GLG 68 78 49 80 26 70 25
GGL 4 7 6 7 24 2
LLG 5 2 11 1 3 5 33
LGL 1 1 2 1 4 1
GLL 4 2 16 1 31 12 2
LLL 1 1 3 4 1
true model: GGL
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 11 2 12 11
LGG 6 9 12 9 1 25 2
GLG 5 8 5 8 24 1 2
GGL 69 80 50 81 26 26 70
LLG 1 2 1 2 3
LGL 4 2 11 1 3 32 5
GLL 4 2 16 1 31 2 12
LLL 1 2 3 1 5



true model: LLG
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 3 1 5
LGG 70 73 63 75 72 7 31
GLG 6 6 11 8 1 32 25
GGL 3 3 5 4 1 5 1
LLG 17 17 15 13 21 14 35
LGL 1 2 1 9 1
GLL 1 4 2 21 2
LLL 1 11 1
true model: LGL
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 3 1 5
LGG 71 73 65 75 73 31 8
GLG 3 3 5 4 1 1 5
GGL 6 6 10 8 1 26 34
LLG 1 1 1 1 8
LGL 17 17 15 13 20 34 13
GLL 1 3 2 1 20
LLL 1 1 11
true model: GLL
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 6 1 11 1 1
LGG 6 9 10 9 2 2
GLG 13 18 10 20 25 13 2
GGL 13 18 10 20 26 2 13
LLG 3 2 4 1 1 10 4
LGL 3 2 5 1 1 4 11
GLL 53 51 52 49 33 57 58
LLL 3 1 8 3 10 10


true model: GGG
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 72 8 6 6 7
LGG 7 28 29 28 3 23 24
GLG 8 36 22 37 22 4 24
GGL 8 33 22 34 23 23 3
LLG 1 1 5 5 3 33
LGL 2 1 5 1 5 36 2
GLL 1 1 8 32 4 4
LLL 1 4 2 3
true model: LGG
model selection metric
LOOCV GCV SCV AIC B0 B1 B2
model selected
GGG 4 1 4 4
LGG 87 90 82 89 94 32 33
GLG 3 4 6 5 1 3 24
GGL 3 4 6 5 1 24 4
LLG 1 1 2 1 1 33
LGL 1 2 1 34 1
GLL 1 1 1 1
LLL 1 1

## true model: GLG

model selection metric

| model selected | LOOCV | GCV | SCV | AIC | B0 | B1 | B2 |
|---|---|---|---|---|---|---|---|
| GGG | 11 | | 2 | | 12 | | 11 |
| LGG | 6 | 10 | 12 | 10 | | 2 | 26 |
| GLG | 68 | 78 | 49 | 80 | 26 | 70 | 25 |
| GGL | 4 | 7 | 6 | 7 | 24 | 2 | |
| LLG | 5 | 2 | 11 | 1 | 3 | 5 | 33 |
| LGL | 1 | 1 | 2 | | 1 | 4 | 1 |
| GLL | 4 | 2 | 16 | 1 | 31 | 12 | 2 |
| LLL | 1 | | 1 | | 3 | 4 | 1 |

## true model: LLG

model selection metric

| model selected | LOOCV | GCV | SCV | AIC | B0 | B1 | B2 |
|---|---|---|---|---|---|---|---|
| GGG | 3 | | 1 | | | | 5 |
| LGG | 70 | 73 | 63 | 75 | 72 | 7 | 31 |
| GLG | 6 | 6 | 11 | 8 | 1 | 32 | 25 |
| GGL | 3 | 3 | 5 | 4 | 1 | 5 | 1 |
| LLG | 17 | 17 | 15 | 13 | 21 | 14 | 35 |
| LGL | 1 | | 2 | | 1 | 9 | 1 |
| GLL | 1 | | 4 | | 2 | 21 | 2 |
| LLL | | | | | 1 | 11 | 1 |

## true model: GGL

model selection metric

| model selected | LOOCV | GCV | SCV | AIC | B0 | B1 | B2 |
|---|---|---|---|---|---|---|---|
| GGG | 11 | | 2 | | 12 | 11 | |
| LGG | 6 | 9 | 12 | 9 | 1 | 25 | 2 |
| GLG | 5 | 8 | 5 | 8 | 24 | 1 | 2 |
| GGL | 69 | 80 | 50 | 81 | 26 | 26 | 70 |
| LLG | 1 | | 2 | | 1 | 2 | 3 |
| LGL | 4 | 2 | 11 | 1 | 3 | 32 | 5 |
| GLL | 4 | 2 | 16 | 1 | 31 | 2 | 12 |
| LLL | 1 | | 2 | | 3 | 1 | 5 |

## true model: LGL

model selection metric

| model selected | LOOCV | GCV | SCV | AIC | B0 | B1 | B2 |
|---|---|---|---|---|---|---|---|
| GGG | 3 | | 1 | | | 5 | |
| LGG | 71 | 73 | 65 | 75 | 73 | 31 | 8 |
| GLG | 3 | 3 | 5 | 4 | 1 | 1 | 5 |
| GGL | 6 | 6 | 10 | 8 | 1 | 26 | 34 |
| LLG | 1 | | 1 | | 1 | 1 | 8 |
| LGL | 17 | 17 | 15 | 13 | 20 | 34 | 13 |
| GLL | 1 | | 3 | | 2 | 1 | 20 |
| LLL | | | | | 1 | 1 | 11 |

## true model: GLL

model selection metric

| model selected | LOOCV | GCV | SCV | AIC | B0 | B1 | B2 |
|---|---|---|---|---|---|---|---|
| GGG | 6 | | 1 | | 11 | 1 | 1 |
| LGG | 6 | 9 | 10 | 9 | | 2 | 2 |
| GLG | 13 | 18 | 10 | 20 | 25 | 13 | 2 |
| GGL | 13 | 18 | 10 | 20 | 26 | 2 | 13 |
| LLG | 3 | 2 | 4 | 1 | 1 | 10 | 4 |
| LGL | 3 | 2 | 5 | 1 | 1 | 4 | 11 |
| GLL | 53 | 51 | 52 | 49 | 33 | 57 | 58 |
| LLL | 3 | 1 | 8 | | 3 | 10 | 10 |

## true model: LLL

model selection metric

| model selected | LOOCV | GCV | SCV | AIC | B0 | B1 | B2 |
|---|---|---|---|---|---|---|---|
| GGG | 2 | | | | 1 | 1 | |
| LGG | 63 | 65 | 55 | 69 | 61 | 8 | 8 |
| GLG | 5 | 5 | 7 | 7 | 2 | 16 | 6 |
| GGL | 4 | 5 | 7 | 6 | 2 | 5 | 17 |
| LLG | 8 | 8 | 8 | 5 | 11 | 14 | 10 |
| LGL | 8 | 8 | 8 | 6 | 11 | 9 | 14 |
| GLL | 1 | 1 | 8 | | 2 | 31 | 30 |
| LLL | 9 | 9 | 6 | 6 | 12 | 15 | 15 |

# Some Conclusions and Questions

LOOCV is pretty good!

# Some Conclusions and Questions

LOOCV is pretty good!

Most accurate coefficient estimates are not necessarily from correct models.

# Some Conclusions and Questions

LOOCV is pretty good!

Most accurate coefficient estimates are not necessarily from correct models.

Are larger variances in some coefficients driving the results?

# Some Conclusions and Questions

LOOCV is pretty good!

Most accurate coefficient estimates are not necessarily from correct models.

Are larger variances in some coefficients driving the results?

Are coefficient estimates equally accurate across metrics?

# Some Conclusions and Questions

LOOCV is pretty good!

Most accurate coefficient estimates are not necessarily from correct models.

Are larger variances in some coefficients driving the results?

Are coefficient estimates equally accurate across metrics?

What happens, ceteris paribus, with greater error variance?