Modeling Monotonic Effects of Ordinal Predictors in Regression Models

Paul Bürkner & Emmanuel Charpentier

Linear Regression

Assume that the predictor term η is a linear combination of the predictor variables multiplied by the regression coefficents:

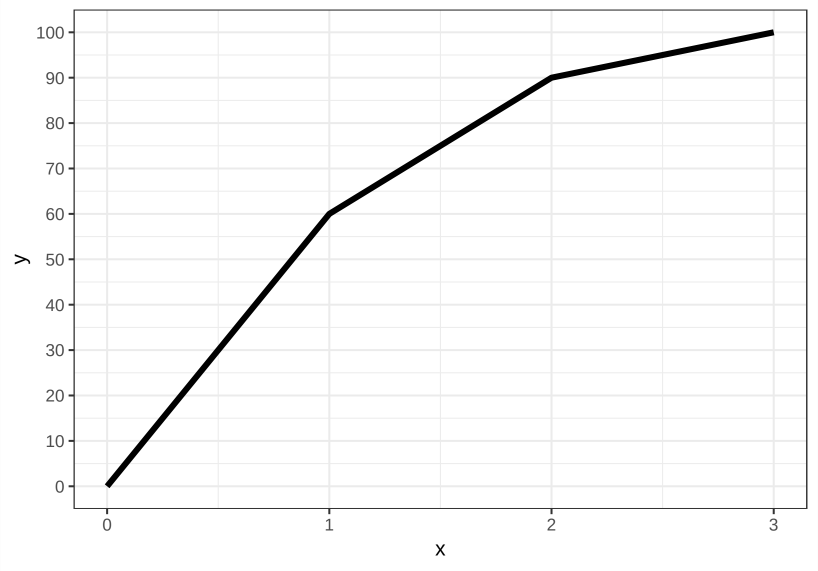
$$\eta = b_0 + \sum_{k=1}^{K} b_k x_k$$

Predictors x_k may be

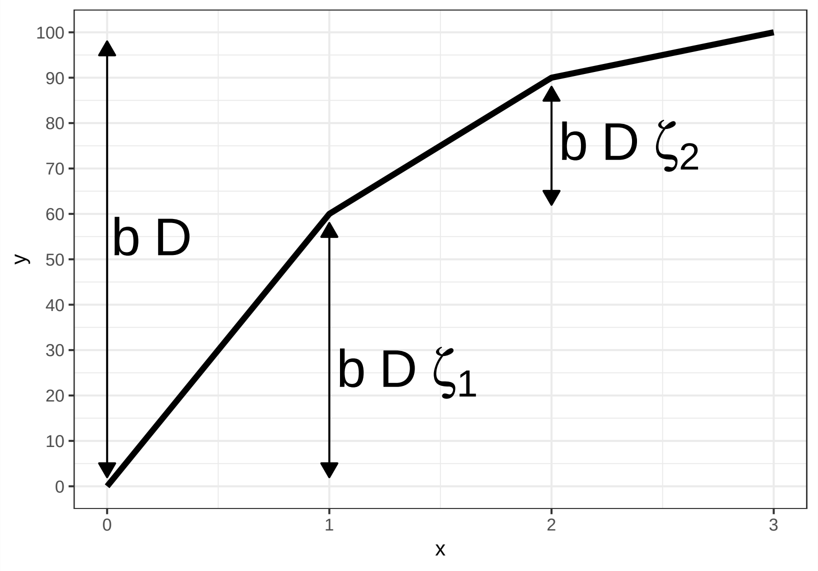
- continuous variables
- coded categorical variables

What about ordinal predictors?

Monotonic Effects: Idea



Monotonic Effects: Idea



Monotonic Effects: Mathematical Formulation

Monotonic regression of an ordinal predictor $x \in \{0, ..., D\}$:

$$\eta = b_0 + bD \sum_{i=1}^{x} \zeta_i$$

- Parameter ζ is a simplex: $\zeta_i \in [0, 1]$ and $\sum_{i=1}^{D} \zeta_i = 1$
- Parameter b may be any real value

Define the monotonic transform:

$$\text{mo}(x, \zeta) = D \sum_{i=1}^{x} \zeta_i$$

Monotonic Effects: Interactions

Ordinary Regression model including the interaction of z and x:

$$\eta = b_0 + b_1\, z + b_2\, x + b_3\, z\, x$$

Generalize to monotonic effects by replacing x with $\mathrm{mo}(x, \zeta)$:

$$\eta = b_0 + b_1\, z + b_2\, \mathrm{mo}(x, \zeta_{b_2}) + b_3\, z\, \mathrm{mo}(x, \zeta_{b_3})$$

Monotonic Effects in a Bayesian Framework

"If you quantify uncertainty with probability you are a Bayesian."

Michael Betancourt

Bayes Theorem:

$$p(\theta \mid y) = \frac{p(y \mid \theta) \, p(\theta)}{p(y)}$$

The monotonic parameters b and ζ are both part of θ

Priors for Monotonic Effects in a Bayesian Framework

Priors on b:

- Any reasonable prior for regression coefficients
- For instance: $b \sim \mathcal{N}(0, s)$ for a fixed standard deviation s

Prior on ζ:

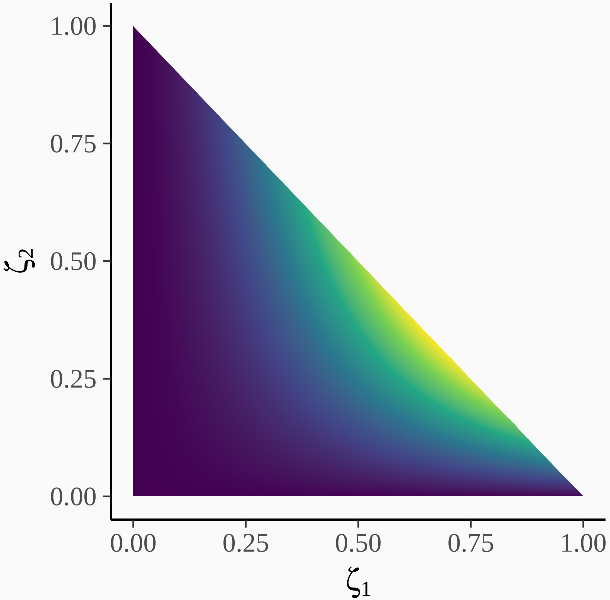
- Dirichlet prior: $\zeta \sim \mathcal{D}(\alpha)$
- α: Concentration parameter of the same length as ζ

Let $\alpha_0 = \sum_{i=1}^{D} \alpha_i$, then:

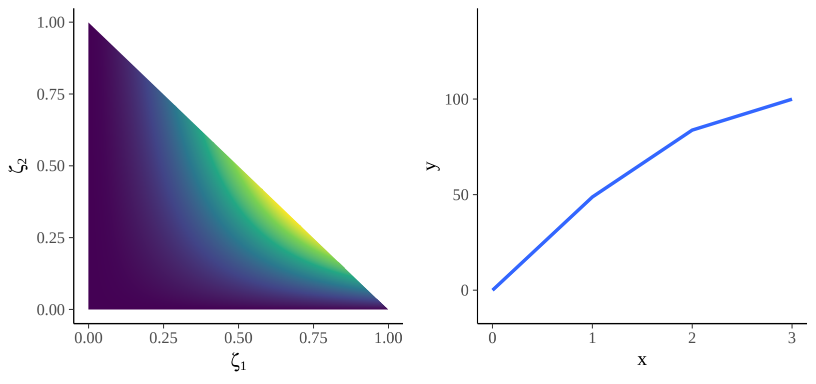
$$\mathbb{E}(\zeta_i) = \frac{\alpha_i}{\alpha_0}$$

$$\mathrm{SD}(\zeta_i) = \sqrt{\frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}}$$

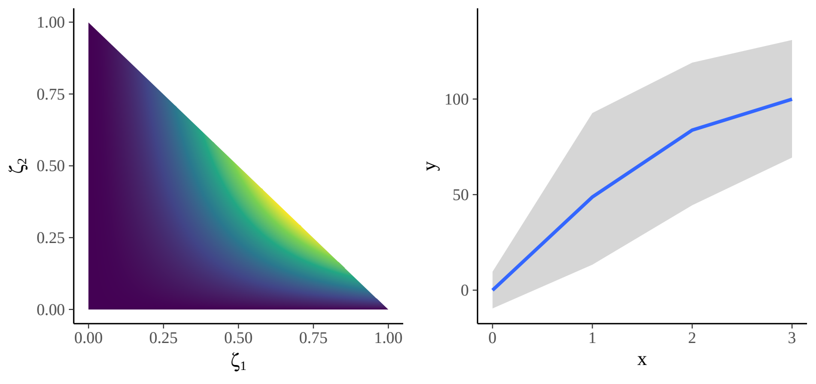
Dirichlet Prior: Visualization for $\alpha = (3, 2, 1)$



Dirichlet Prior: Visualization for $\alpha = (3, 2, 1)$



Dirichlet Prior: Visualization for $\alpha = (3, 2, 1)$



Monotonic effects in the R package brms

Monotonic effect of x on y:

```
y ~ mo(x)
```

Main effects and interaction of x and z:

```
y ~ mo(x) * z
```

Varying effect of x over group g:

```
y ~ mo(x) + (mo(x) | g)
```

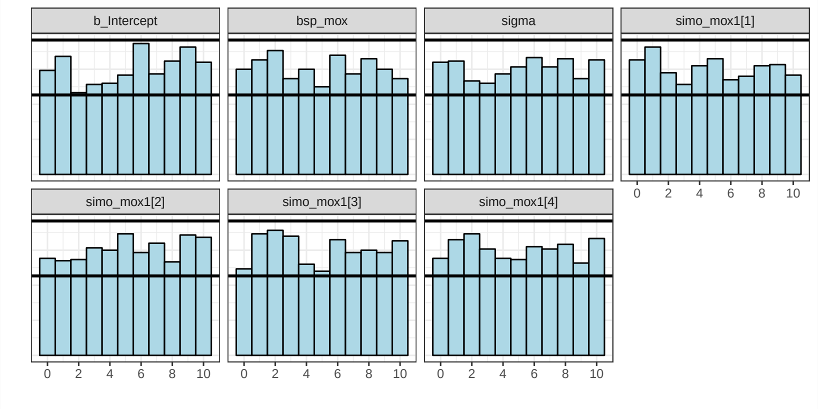
Parameter Recovery

How well can a model recover its own parameters?

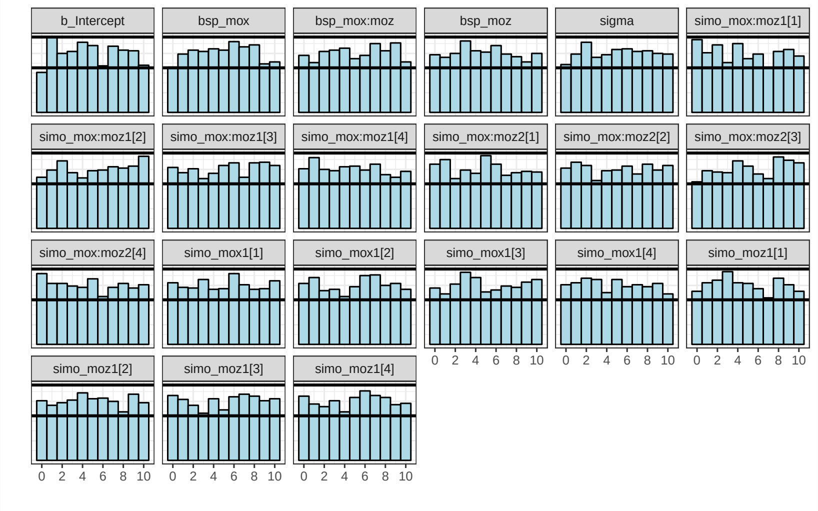
- Simulate data from the model with known parameters
- Fit the model to the simulated data
- Compare estimates to the known parameters

Bayesian version: Simulation Based Calibration (SBC) by Talts, Betancourt, Simpson, Vehtari, & Gelman (2018)

Parameter Recovery: Monotonic Main Effects



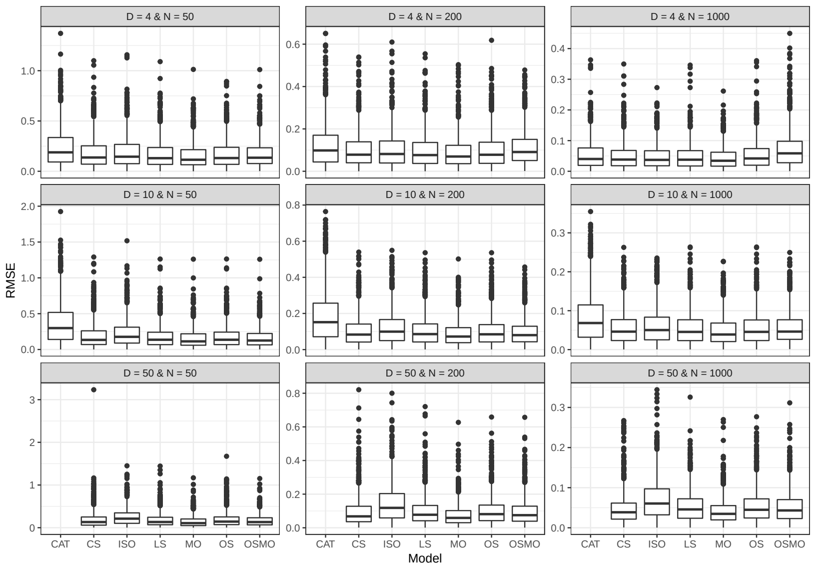
Parameter Recovery: Monotonic Interactions



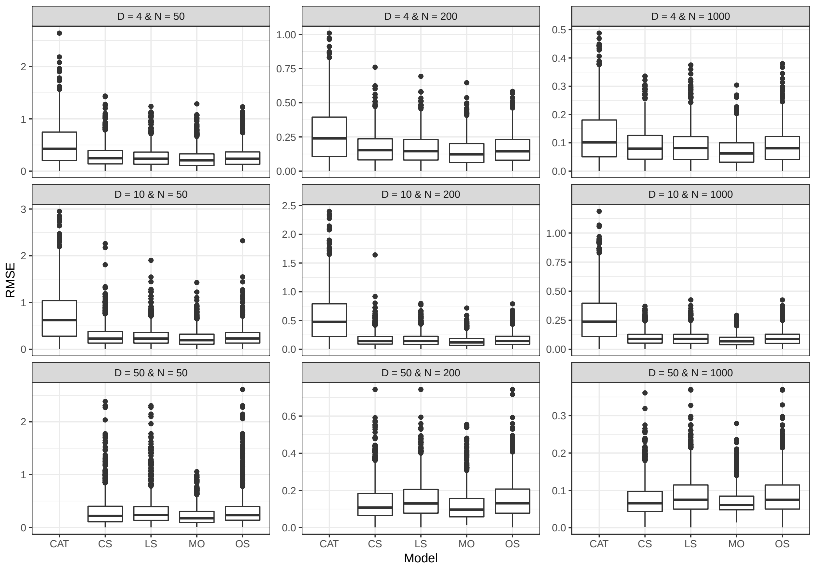
Other Approaches for Modeling Ordinal Predictors

- Continuous linear regression
- Categorical linear regression
- Categorical isotonic regression
- Penalized categorical regression
- Monotonic penalized categorical regression
- Regression splines
- ...

Model Comparison: Monotonic Main Effects



Model Comparison: Monotonic Interactions



Summary

Thank you!

References

Bürkner, P., & Charpentier, E. (in review). Modeling Monotonic Effects of Ordinal Predictors in Regression Models. https://psyarxiv.com/9qkhj/

Gertheiss, J., Hogger, S., Oberhauser, C., & Tutz, G. (2011). Selection of ordinally scaled independent variables with applications to international classification of functioning core sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *60*(3), 377–395.

Gertheiss, J., & Tutz, G. (2009). Penalized regression with ordinal predictors. *International Statistical Review*, *77*(3), 345–365.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating bayesian inference algorithms with simulation-based calibration. *arXiv Preprint arXiv:1804.06788*.

Appendix

Case Study: Measures of Chronic Widespread Pain (CWP)

Objective: Predict subjective physical health by measures of CWP

Examples for CWP measures:

- Impairments in walking
- Impairments in moving around

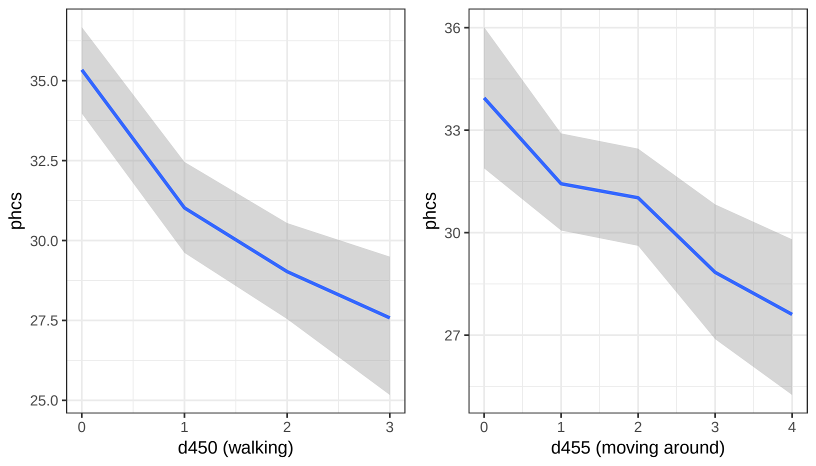
Scale from 0 (‘no problem’) to 4 (‘complete problem’)

Reference: Gertheiss, Hogger, Oberhauser, & Tutz (2011)

Plausible assumption: CWP measures have **monotonic effects**

Case Study: Model Specification

```r
library(brms)
fit1 <- brm(phcs ~ mo(d450) + mo(d455), data = cwp)
```



Simulation Based Calibration

How well can a model recover its own parameters?

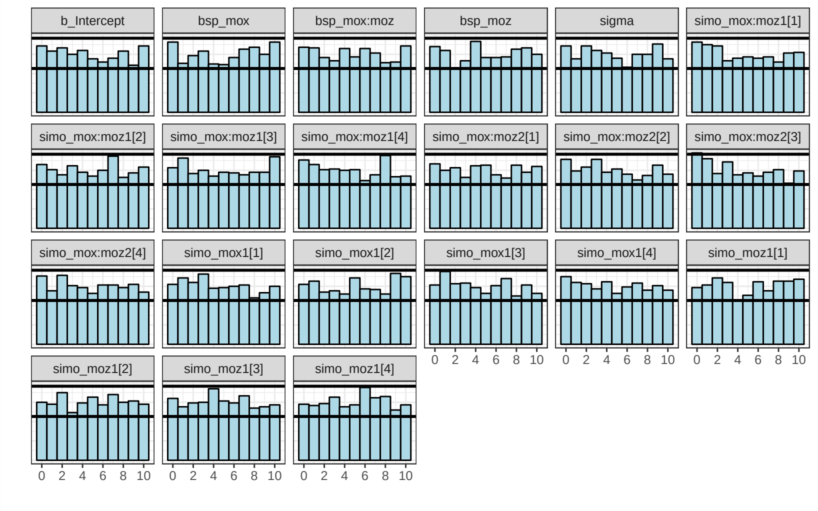
Steps of Simulation Based Calibration (SBC):

- Sample $\tilde{\theta} \sim p(\theta)$ from the prior
- Sample $\tilde{y} \sim p(y \mid \tilde{\theta})$ from the likelihood
- Sample $\{\theta_1, \ldots, \theta_L\} \sim p(\theta|\tilde{y})$ from the posterior
- Compute the rank statistic $r(\{\theta_1, \ldots, \theta_L\} \mid \tilde{\theta})$
- Repeat the process multiple times
- Plot the rank statistics in a histogram

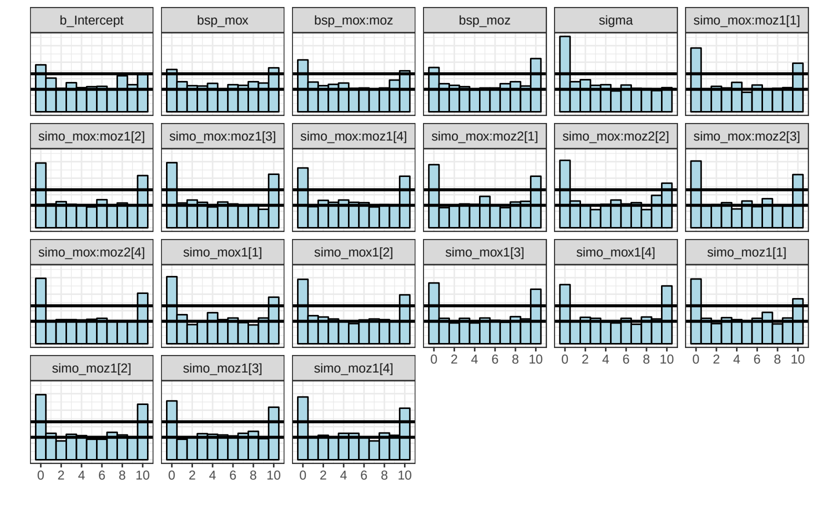
For well callibrated models the histogram is (approximately) uniform

Reference: Talts et al. (2018)

Parameter Recovery: Interactions with many Predictor Categories (1)



Parameter Recovery: Interactions with many Predictor Categories (2)



Other Approaches for Modelling Ordinal Predictors

Categorical isotonic regression:

- Estimate group means of ordinal categories such that $\mu_0 < \mu_1 < ... < \mu_C$
- Equivalent to monotonic effects in simple cases
- Harder to penalize via priors

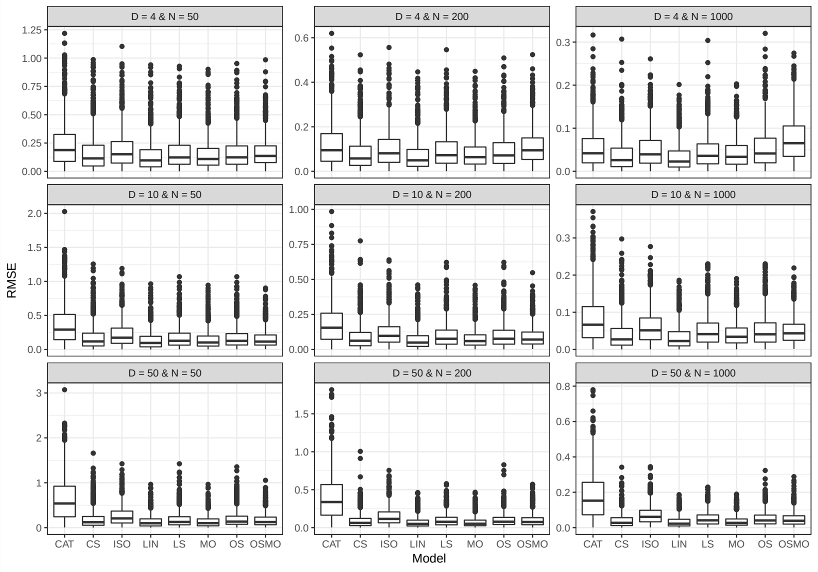
Penalized regression (Gertheiss & Tutz, 2009):

- Apply dummy coding on the ordinal variable
- Penalize larger differences between adjacent categories via

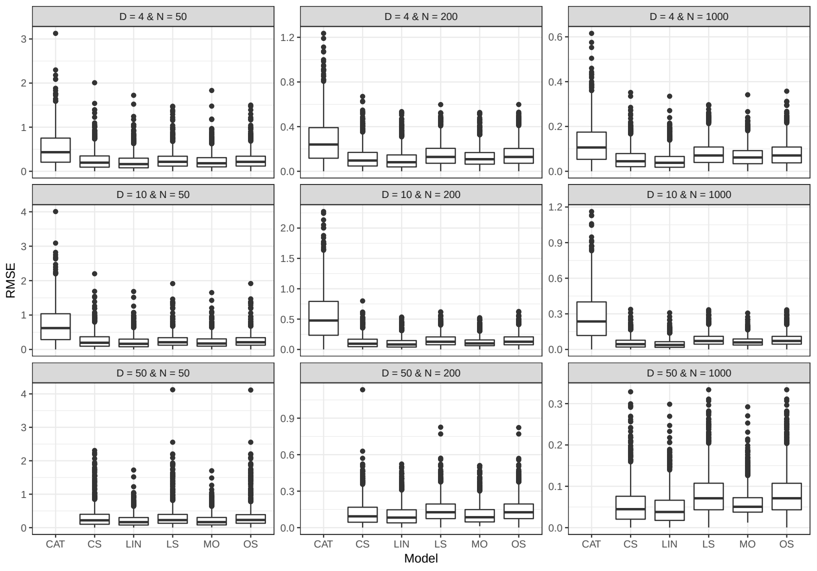
$$J(b) = \sum_{i=1}^{D} (b_i - b_{i-1})^2$$

- Closely related to regression splines
- No monotonicity constraint

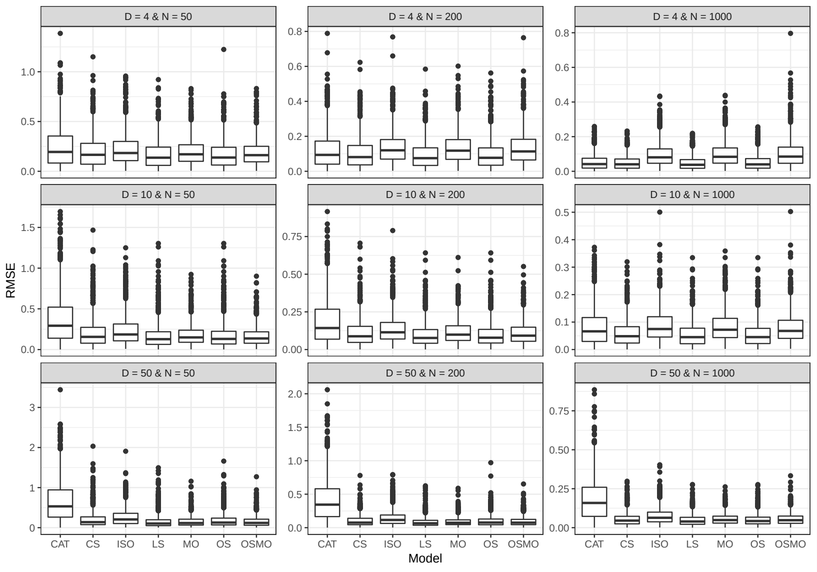
Model Comparison: Linear Main Effects



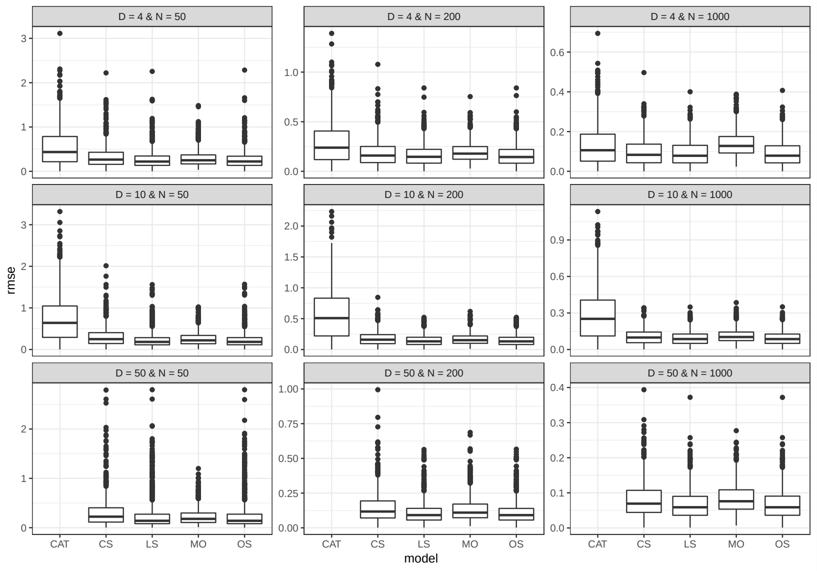
Model Comparison: Linear Interactions



Model Comparison: Categorical Main Effects



Model Comparison: Categorical Interactions



Proof: Monotonicity

Proposition: Monotonic effects are indeed monotonic.

Proof idea:

$$b\text{mo}(x+1,\zeta) - b\text{mo}(x,\zeta) = bD\sum_{i=1}^{x+1}\zeta_i - bD\sum_{i=1}^{x}\zeta_i = bD\zeta_{x+1}$$

Since $D > 0$ and $\zeta_{x+1} > 0$, the linear predictor is monotonically increasing if $b \geq 0$ and monotonically decreasing if $b \leq 0$.

Proof: Conditional Monotonicity

Proposition: If all ζ belonging to x are the same, then the predictions are monotonic in x conditional on all possible values of all other predictors.

Proof idea:

$$\eta(x) = b_0 + \sum_{k=1}^{K} b_k D_k \sum_{i=1}^{x} \zeta_i = b_0 + \left(\sum_{k=1}^{K} b_k D_k\right)\left(\sum_{i=1}^{x} \zeta_i\right)$$

If we define $b = \sum_{i=1}^{K} b_i D_i$ we see that $\eta(x)$ is monotonic in x with the sign of the effect determined by the sign of b.

Counter Example to General Conditional Monotonicity

Model: $\eta = b_0 + b_1\, z + b_2\, \mathrm{mo}(x, \zeta_{b_2}) + b_3\, z\, \mathrm{mo}(x, \zeta_{b_3})$