

Can predicting data improve model interpretation and inferences? Using Stan to fight the assumption of independence.

František Bartoš¹, Jiří Lukavský², Filip Děchtěrenko^{1,2}

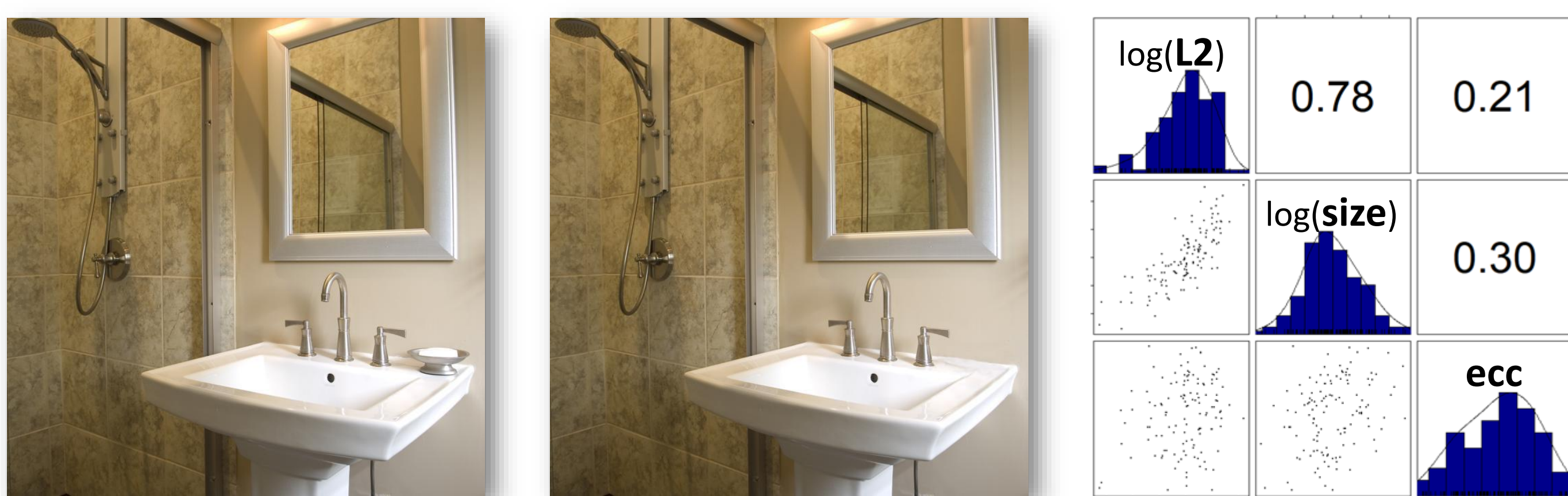
¹Faculty of Arts, Charles University, Prague, Czech Republic, ²Czech Academy of Sciences, Prague, Czech Republic

Goal:

- model interpretation with highly correlated predictors

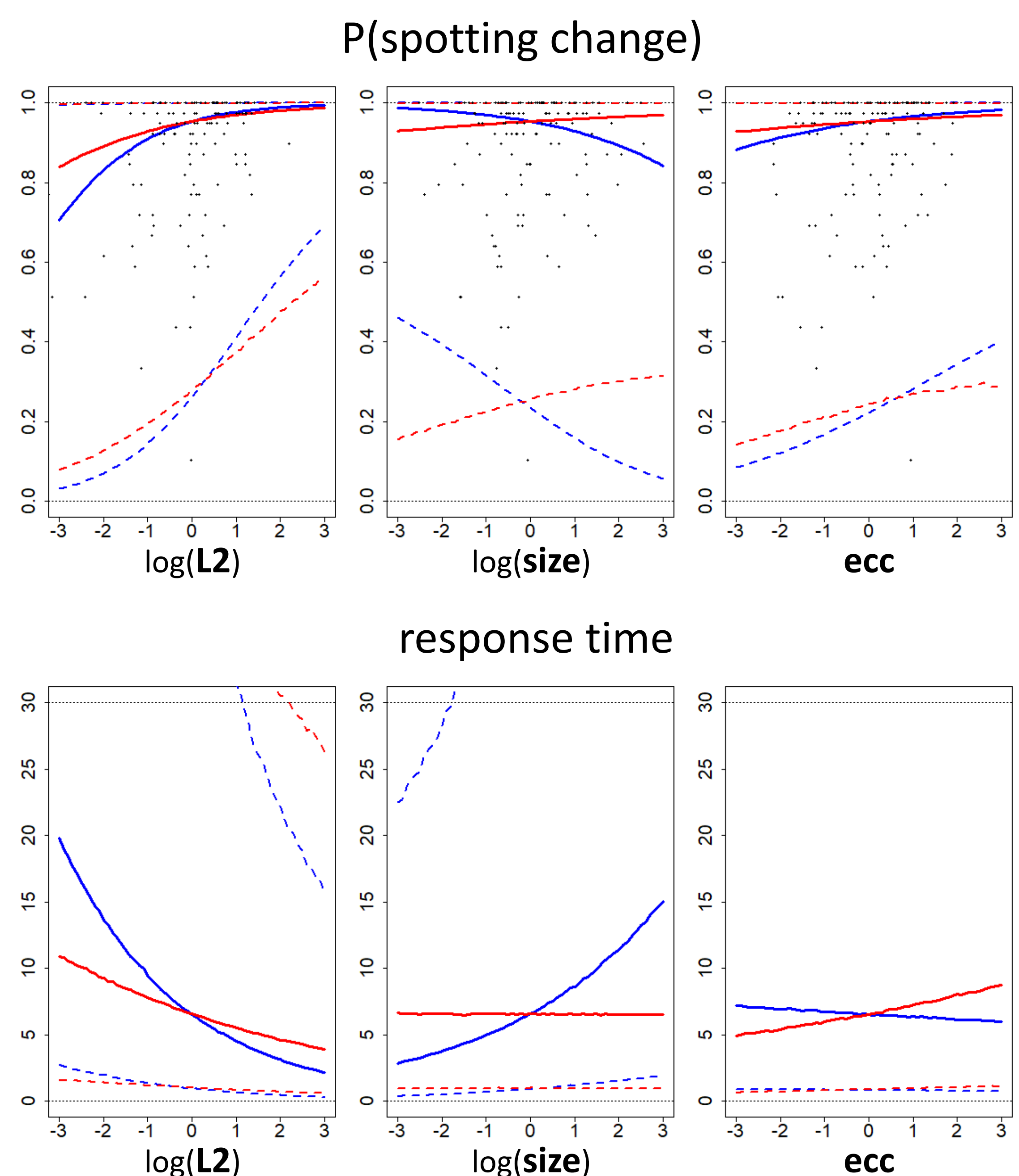
Dataset:

- cognitive psychology experiment
- participants asked to indicate, whether are two interchanging images the same
- 40 participants
- 128 pairs of images belonging to 8 categories
- trial features:
 - **size** of the change
 - **eccentricity** of the change
 - measure of dissimilarity of the two images in CNN using **L2**



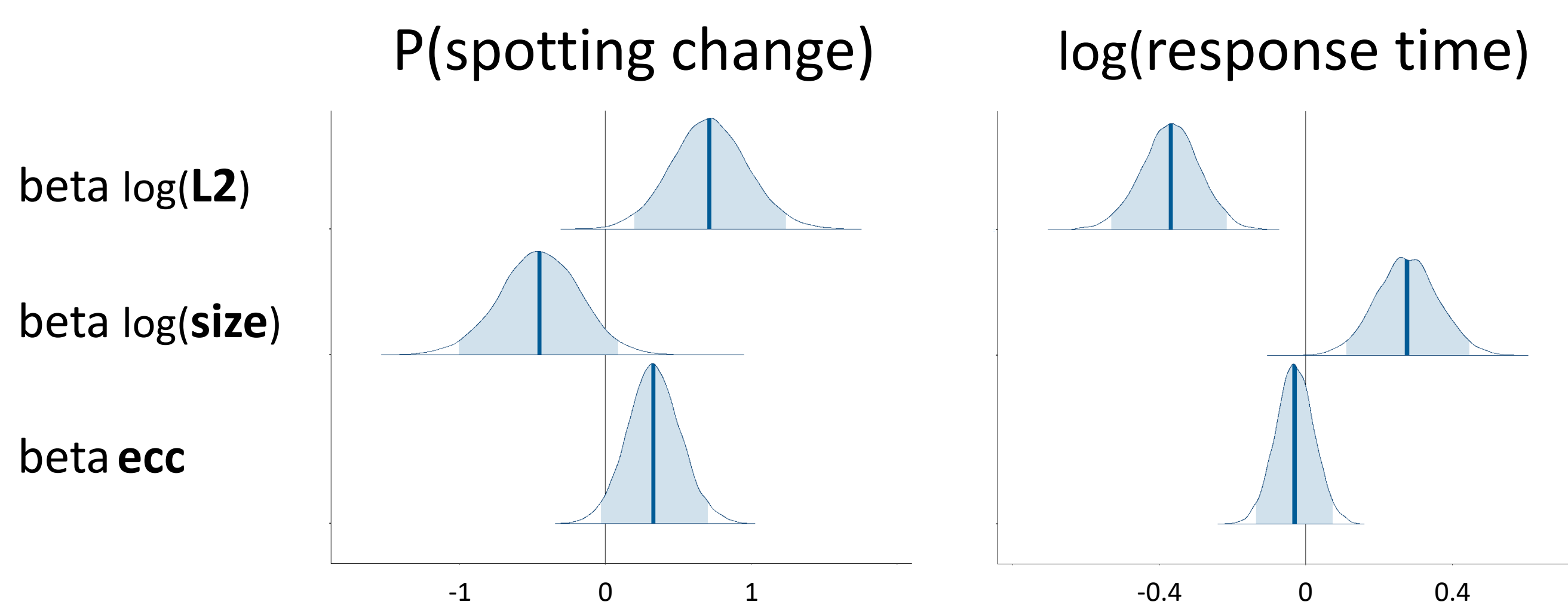
Interpreting the results with partial dependence plots

- estimating the grand mean and expected range for 95% of future observations
- classical partial dependence plots in **blue**
- partial dependence plots accounting for correlation in **red**



Models:

- two hierarchical models in Stan
- logistic regression for the probability of spotting the change
- linear regression with censoring for the response times



Paradox: the larger the size of the change, the lower the probability of spotting it and the longer the response times?



Interested in the models?

Check them out in the ShinyStan application!

http://bit.ly/stan_models



Acknowledgements:

This project was partially supported by Czech Science Foundation (16-07983S) and by student grant GA UK no. 604218

Construction of the partial dependence plots accounting for correlation between predictors

- 1) select a range of values for the predictor of interest
- 2) estimate expected values of the other predictors in the selected range
- 3) input the values of the predictor of interest and estimated values of the other predictors to the model for the response variable

Summary:

- Effects of correlated predictors are often hard to interpret
- We propose a solution how to construct partial dependence plots accounting for the correlation
- Contribution of each predictor is easier to understand and the method facilitates more accurate interpretation of results

What is the impact of different correlations between predictors on partial dependence plots?

Try it yourself in the interactive Shiny application!

http://bit.ly/predicting_data



Contact: email@frantisek-bartos.info / @BartosFra