Performance of COR model on the ABC data sets

Paula Parpart
Department of Experimental Psychology, University College London

November 10, 2015

1 Simulation

The model was fit to the classic 20 abc data sets from Czerlinski et al., 1999. In these classic data sets, the attributes are binarised (at the median) into 0 and 1 (from originally continuous data). We create all possible pairwise comparisons which ends up in attribute data containing the possible values 0,1 or -1. Number of attributes vary between data sets. The dependent variable is always binary and coded as -1 and +1. The COR model is cross-validated on each data set by splitting the total number of pairwise comparisons randomly into training and test set. The size of the training set is varied between 5%, 10%, 30%, 50%, 70%, and 90% of all instances, and the test set represents the complementary set of instances always. For each training set size, the cross-validation split into training and test set is repeated k=100 times and performance is averaged across all of them. Plots below demonstrate the generalization performance of the COR model for a range of penalization parameters $\theta = [0, 700]$, and as a function of the training set size.

This updated version of the document takes care of the mcmc convergence issue due to a bug in the original mcmc code, which results in faster mcmc convergence. Additionally, the number of train-test set separations was increased from k=10 to k=100 which decreased the error bars.

The decision rule plotted here is the TTB decision rule as we want to compare performance between TTB Heuristic and Linear Regression as well as strategies in between. Because the attribute data is binary in this case, we can apply the TTB decision rule which was not possible in the other real-world data sets with continuous attributes.

$$\hat{y} = \mathrm{sgn}\left(y_{\arg\max\limits_{j} |y_{ij}|} \right)$$

Guessing: Whether or not the TTB decision rule allows for guessing in non-discriminating cases makes a difference for the overall generalization performance. As the definition of the TTB heuristic is "when no cues discriminate between the two alternatives, make a guess", our TTB decision rule should probably also allow for guessing in these very common non-discriminating cases. However, when the decision rules involve the guessing option, results are more noisy and it becomes harder to find agreement with the actual TTB heuristic in the limit (e.g., if half of the test set are guessing cases then performance can differ a lot). So, in order to be able to establish convergence with the heuristics, I did not include the guessing option below, and whenever a decision rule predicts indifference (0), it now predicts zero. I applied the same rule to the TTB Heuristic and Linear

Regression. This leads to overall slightly lower performance as the dependent variable only contains -1 and +1.

1.1 House

Number of pairwise comparisons: 231
Number of attributes: 10
Class variable: Which house had the higher actual sale price? (+1,-1)
Average (absolute) correlation between attributes: 0.35
Minimum correlation between attributes: -0.27
Maximum correlation between attributes: 0.91

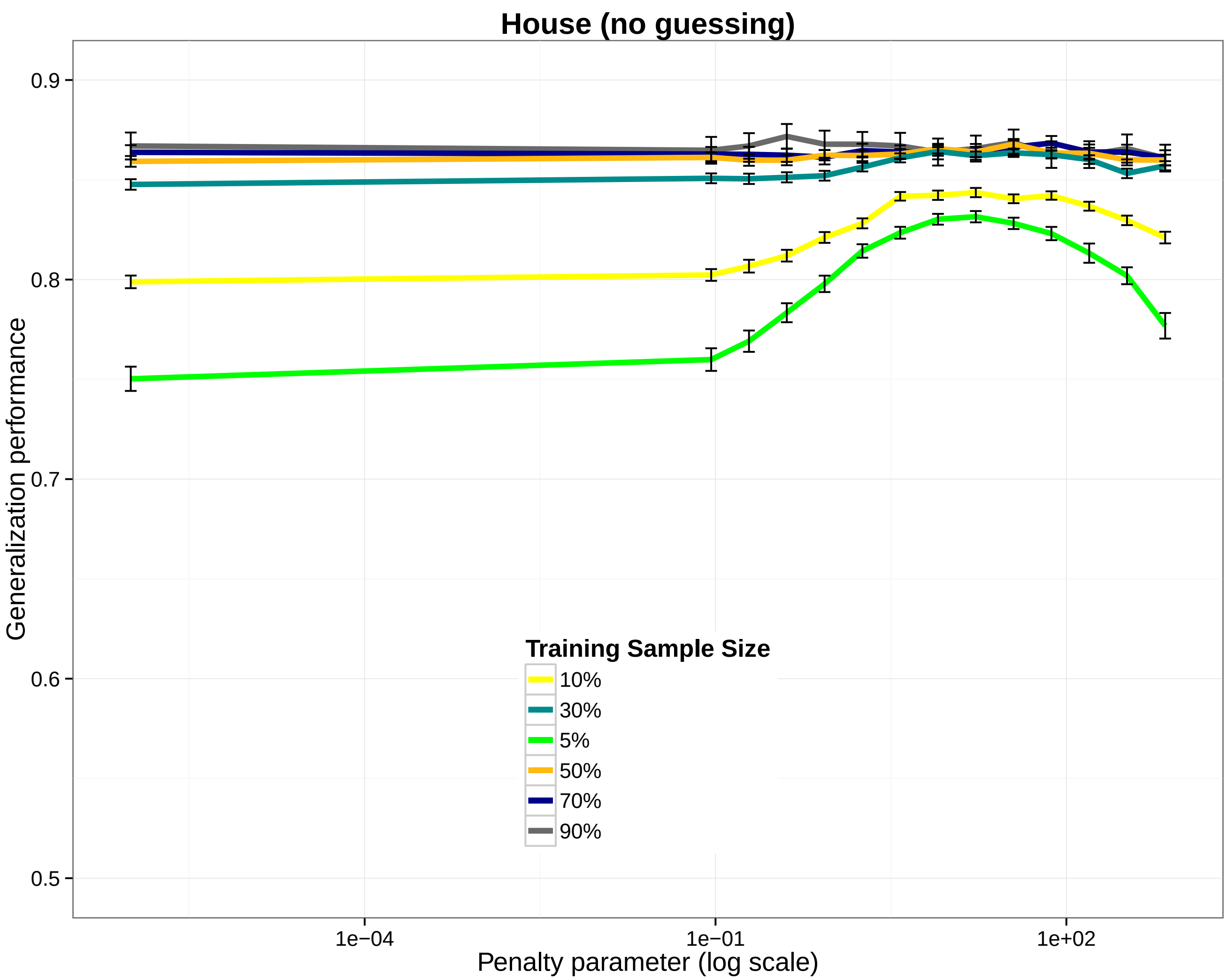

Figure 1: Performance of the COR model as a function of the penalization parameter. Training sample size varied between 5%, 10%, 30%, 50%, 70%, and 90% of all instances.

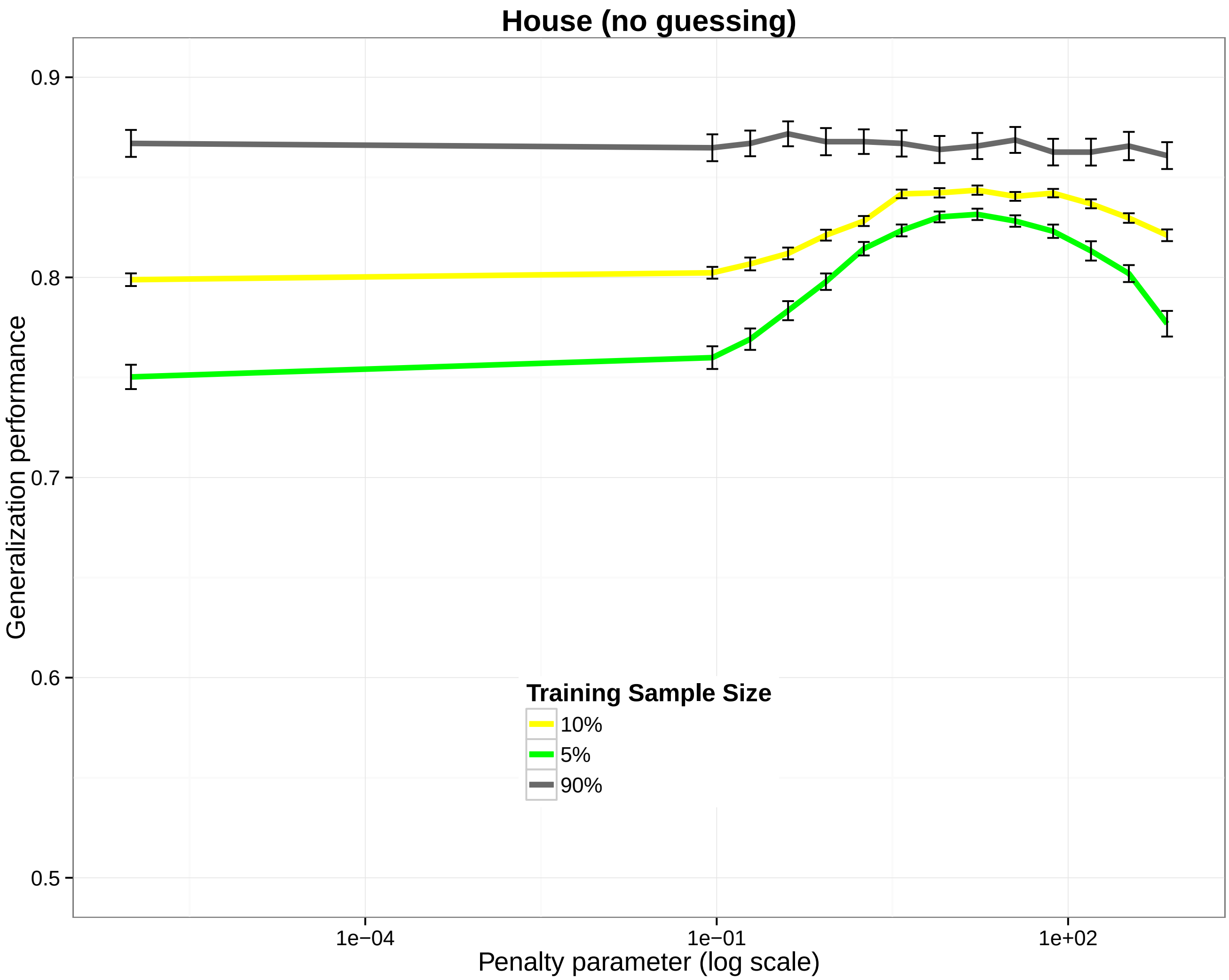

Figure 2: Performance of the COR model as a function of the penalization parameter. Training sample size varied between 5%, 10%, 30%, 50%, 70%, and 90% of all instances.

1.2 Ozone

Number of pairwise comparisons: 55
Number of attributes: 3
Class variable: Which case had the higher levels of ozone in San Francisco? (+1,-1)
Average (absolute) correlation between attributes: 0.64
Minimum correlation between attributes: 0.45
Maximum correlation between attributes: 0.83

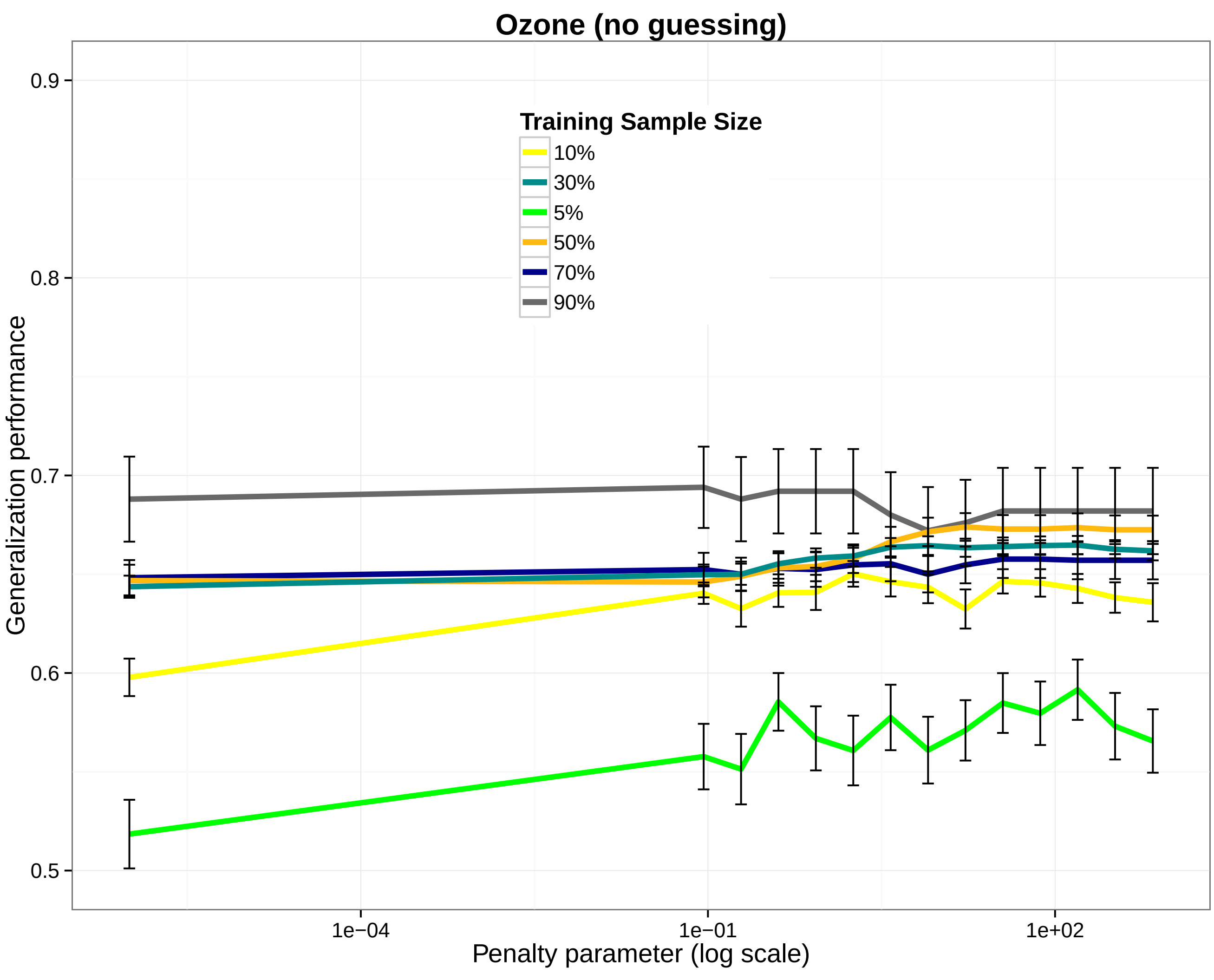

Figure 3: Performance of the COR model as a function of the penalization parameter. Training sample size varied between 5%, 10%, 30%, 50%, 70%, and 90% of all instances.

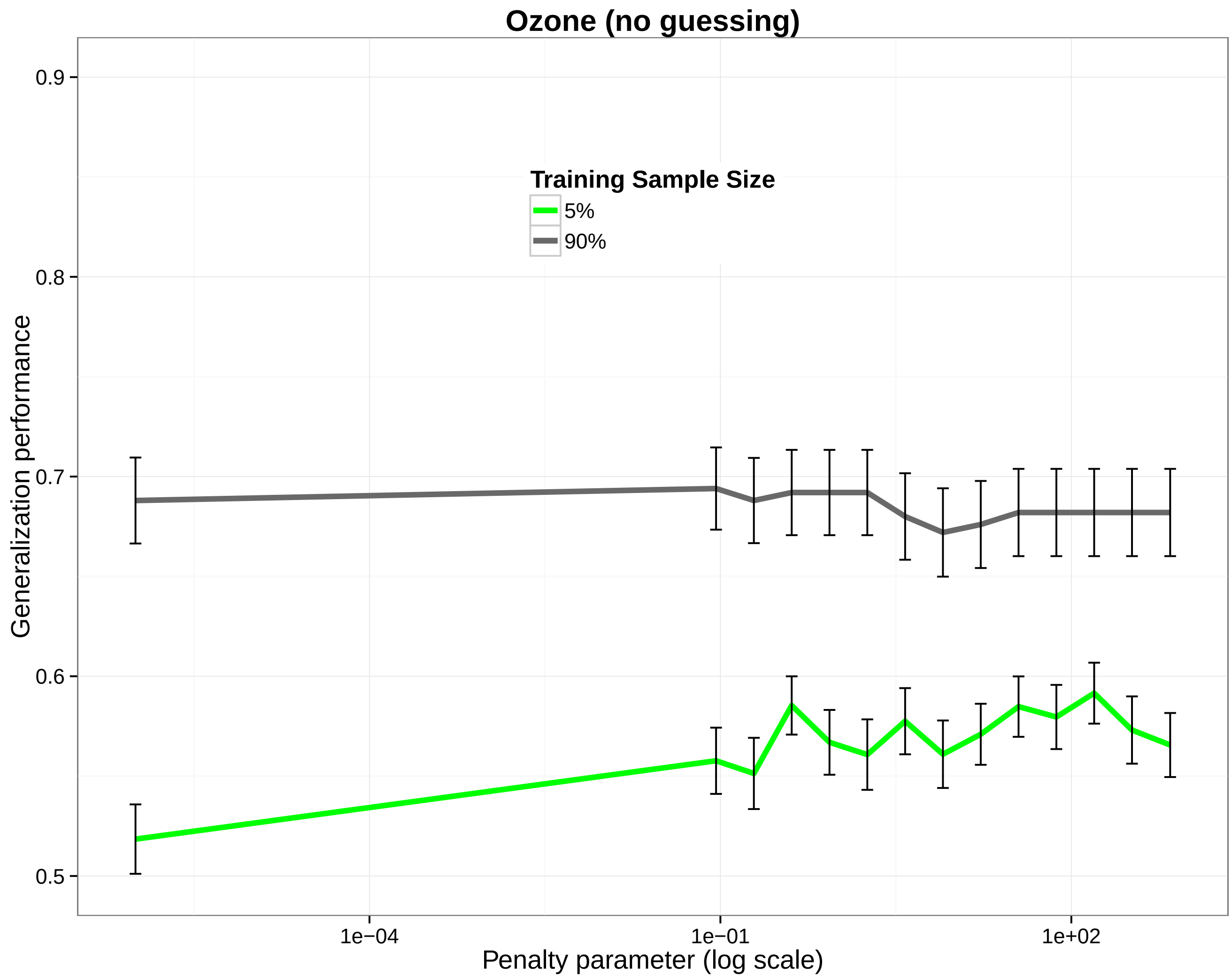

Figure 4: Performance of the COR model as a function of the penalization parameter. Training sample size varied between 5%, 10%, 30%, 50%, 70%, and 90% of all instances.

1.3 Mortality

Number of pairwise comparisons: 190
Number of attributes: 15
Class variable: Which place had the higher mortality rates? (+1,-1)
Average (absolute) correlation between attributes: 0.21
Minimum correlation between attributes: -0.61
Maximum correlation between attributes: 1

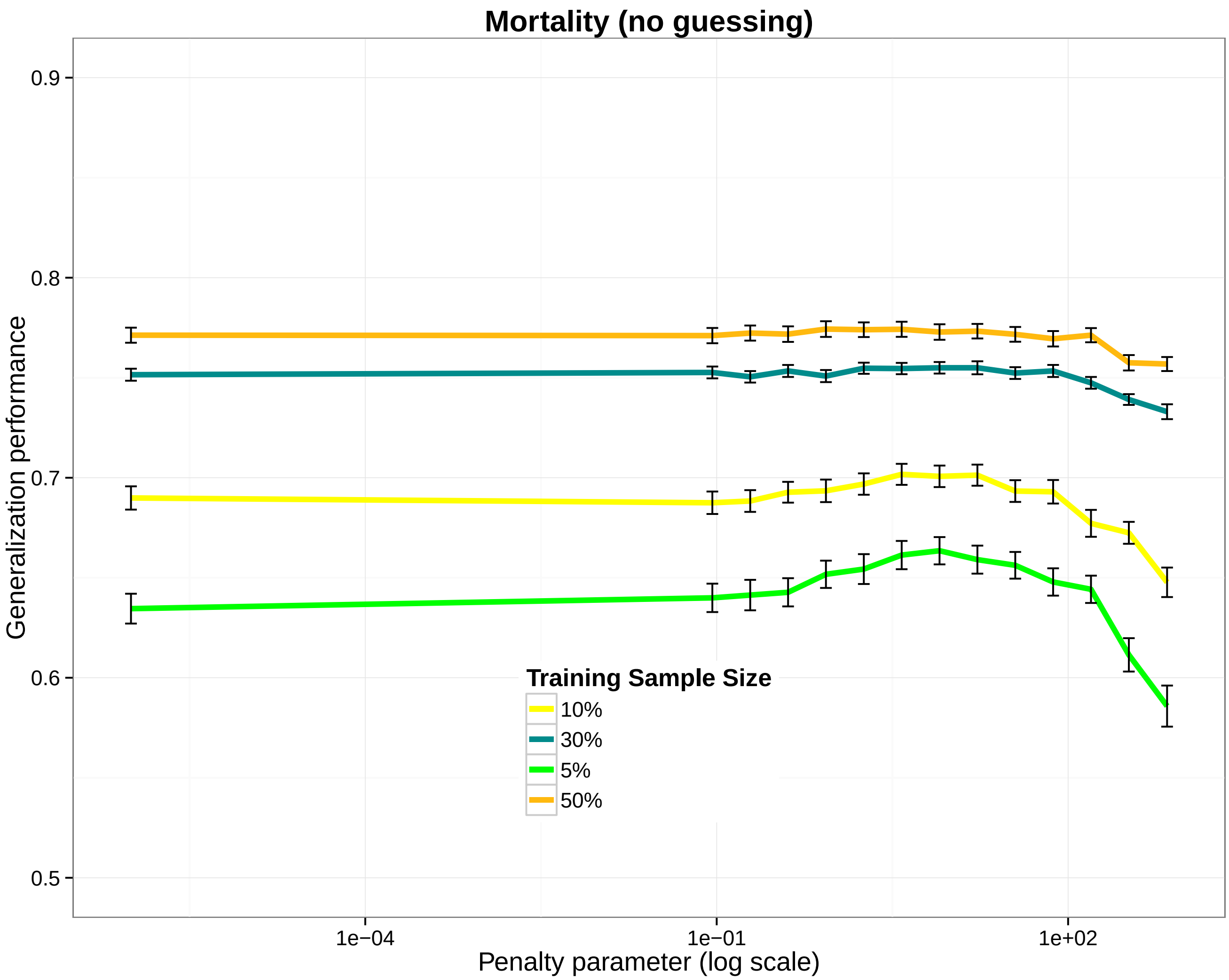

Figure 5: Performance of the COR model as a function of the penalization parameter. Training sample size varied between 5%, 10%, 30%, 50%, 70%, and 90% of all instances.