Simulation of PCB with resampling

Xuelong Wang

2018-09-13

Contents

1   Motivation

Based on the simulation studies, we found the proposed method could work well even if the covariates' distributions are correlated and long tail. More specifically, the proposed method can estimate the main and interactive effect's variance without much bias. Thus, we have some reason to believe that the proposed method could also work well on the PCB data.

We want to conduct the previous simulation studies on PCB data. However, there is a issue to be solved. One of the assumptions of GCTA method is that the covariates are random variables. Thus, the covariates X are randomly sampled for each iteration during the simulation procedure. But it cannot be available for PCB data, because we only have one set of data with 1000 observations.

2   Re-sampling procedure

To solve the problem mentioned above, we could used the resampling PCB's observations without replacement to mimic the randomization of the covariates. Although those re-sampled data is not the true random variable, it more or less can give as an idea that how well the proposed method can work on the PCB data.

2.1   Bootstrap VS Resampling

Resampling method can keep the observed covariance structure of the PCB data. However, since the bootstrap using sampling with replacement, it cannot keep the original structure.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

$$X_{resample} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$X_{bootstrap} = \begin{bmatrix} x_1 \\ x_2 \\ x_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} & \sigma_{22} \\ \sigma_{21} & \sigma_{22} & \sigma_{22} \end{bmatrix}$$

2.2 re-sample size

We need to find an appropriate sample size which is not too large to reduce the randomization or too sample to increase the estimations variation.

$$m = p * n, \text{ where } p = (0.6 - 0.9)$$

3 Simulation result

3.1 Total effect estimation
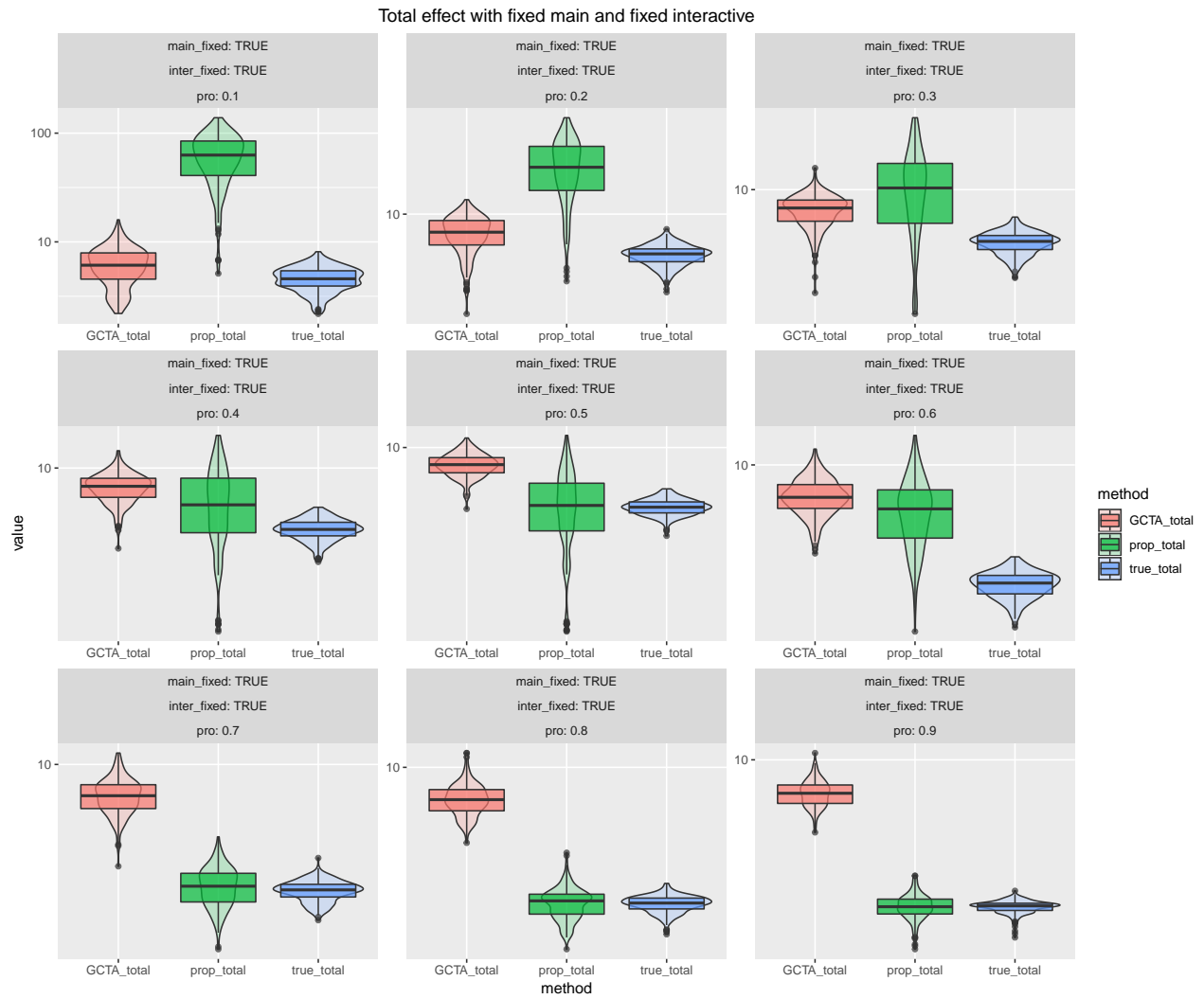
Figure 1: Total effect estimation with fixed main and fixed interaction
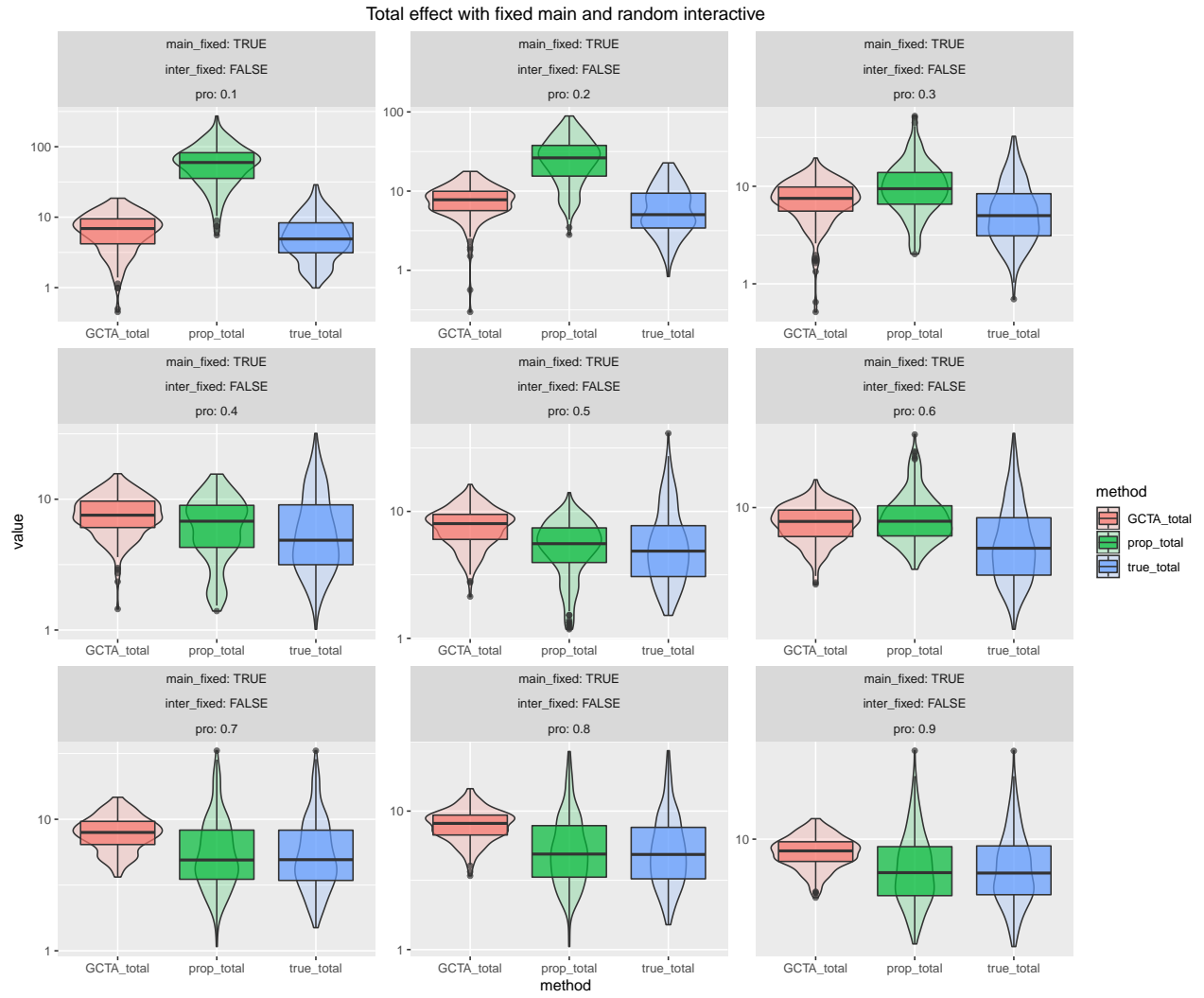
Total effect with fixed main and random interactive

Figure 2: Total effect estimation with fixed main and random interaction
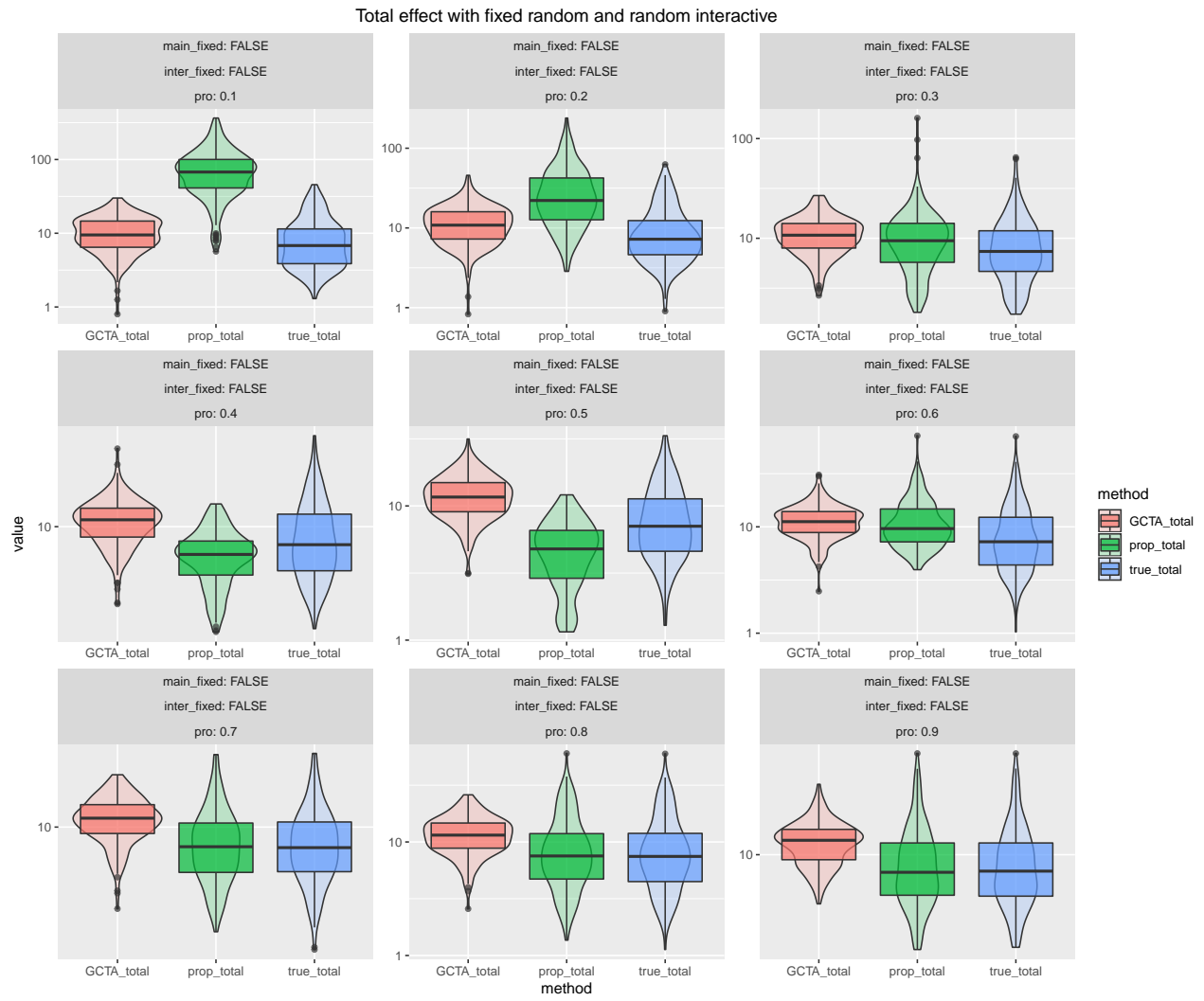
Figure 3: Total effect estimation with random main and random interaction

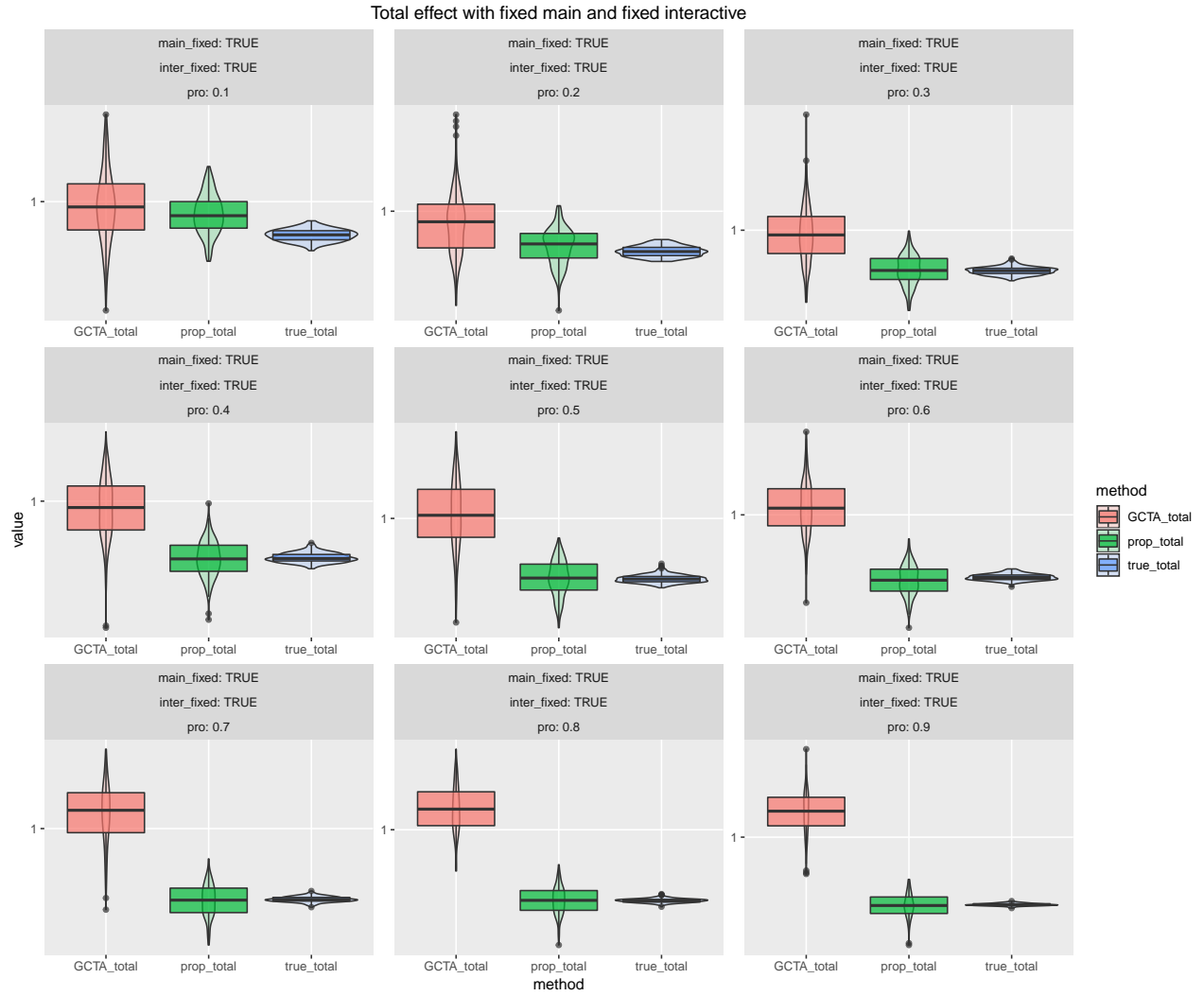## 3.2 Total effect estimation with p = 6

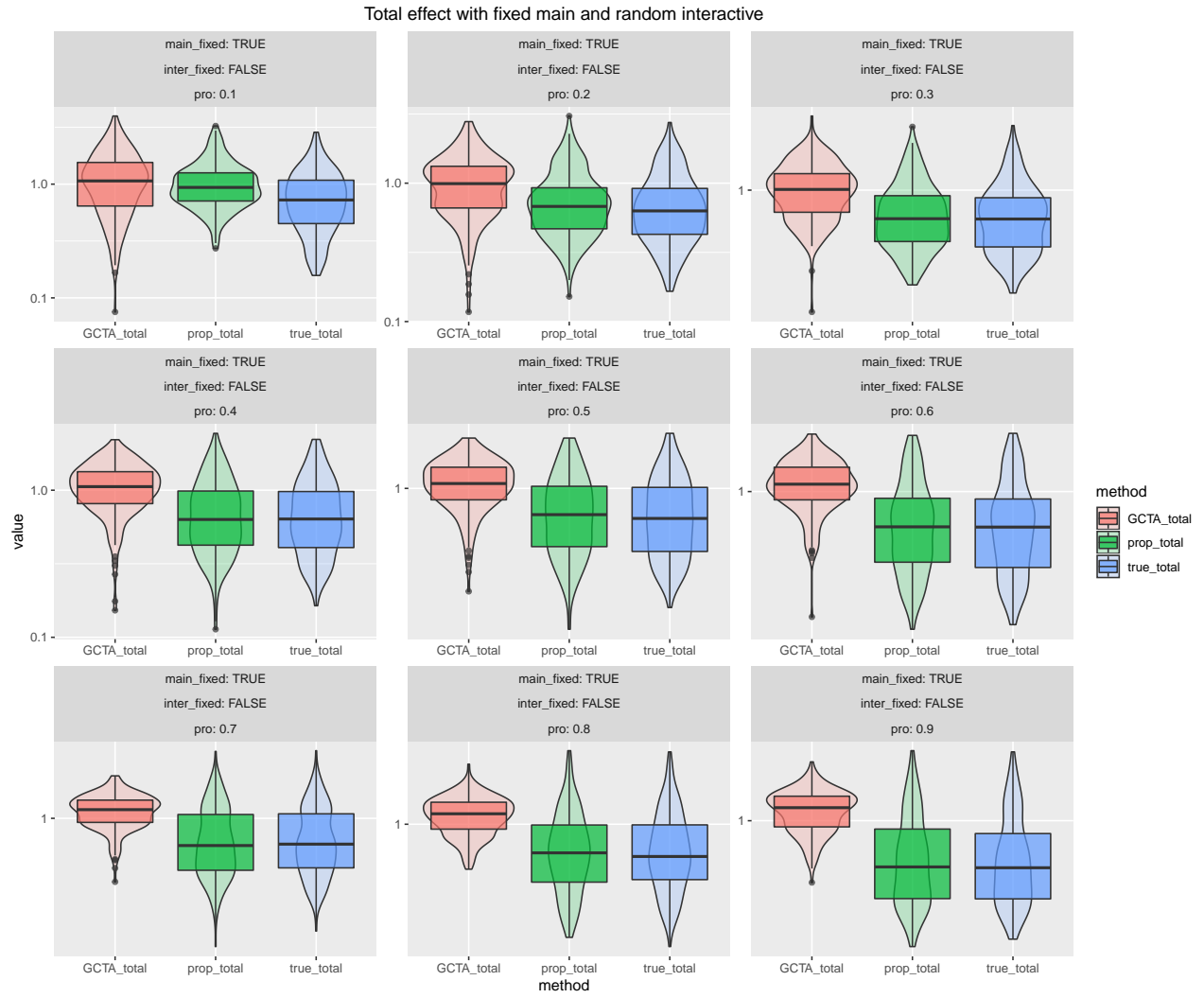Figure 4: Total effect estimation with fixed main and fixed interaction

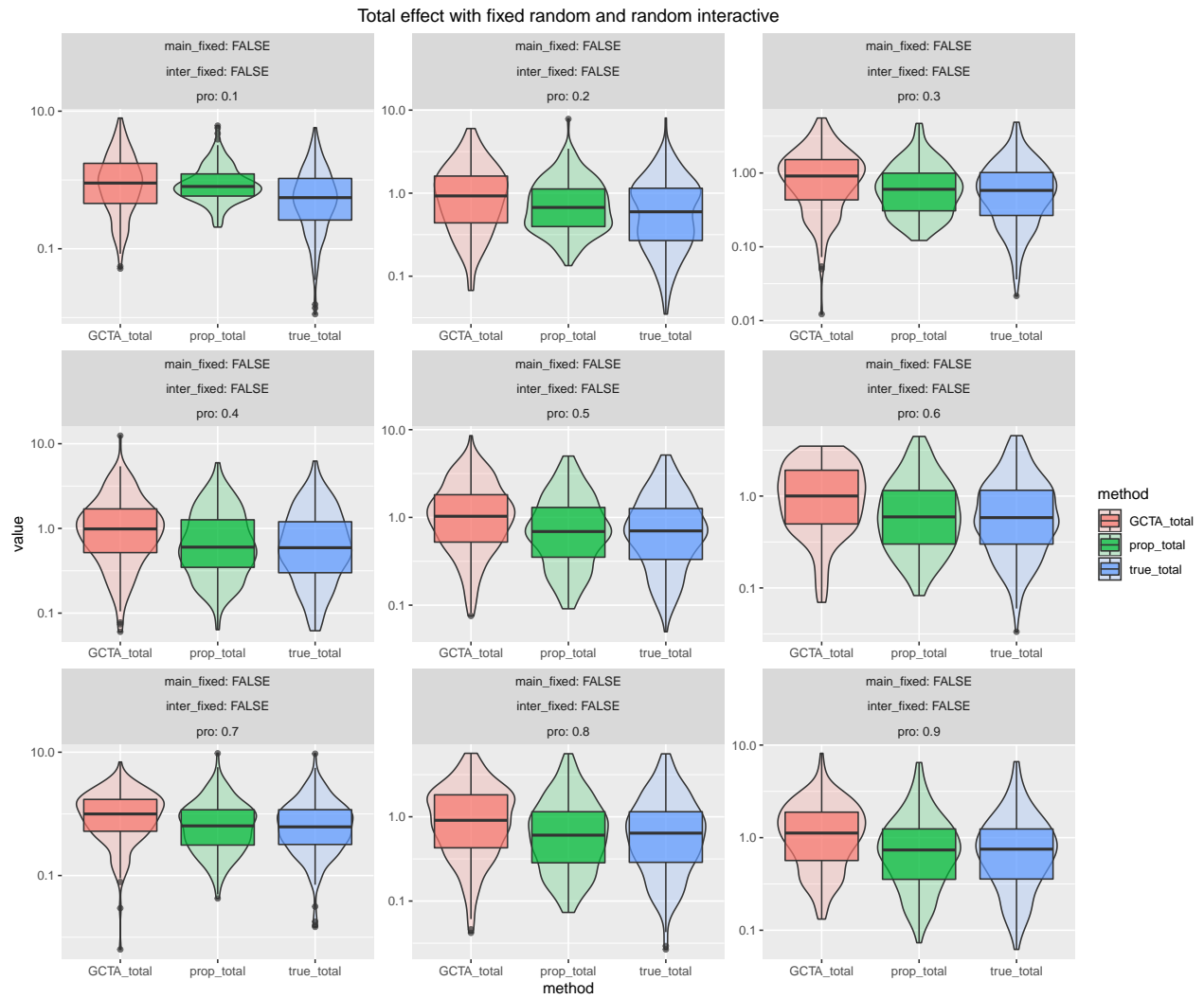Figure 5: Total effect estimation with fixed main and random interaction

Figure 6: Total effect estimation with random main and random interaction

3.3 Main effect estimation (without interaction term in the model)

The purpose of this simulation is try to estimate the main effect only given the true mode has interaction effect. Since we focus on the main effect and ignore the interactive effect, the model we are using to do the estimation is a wrong model.
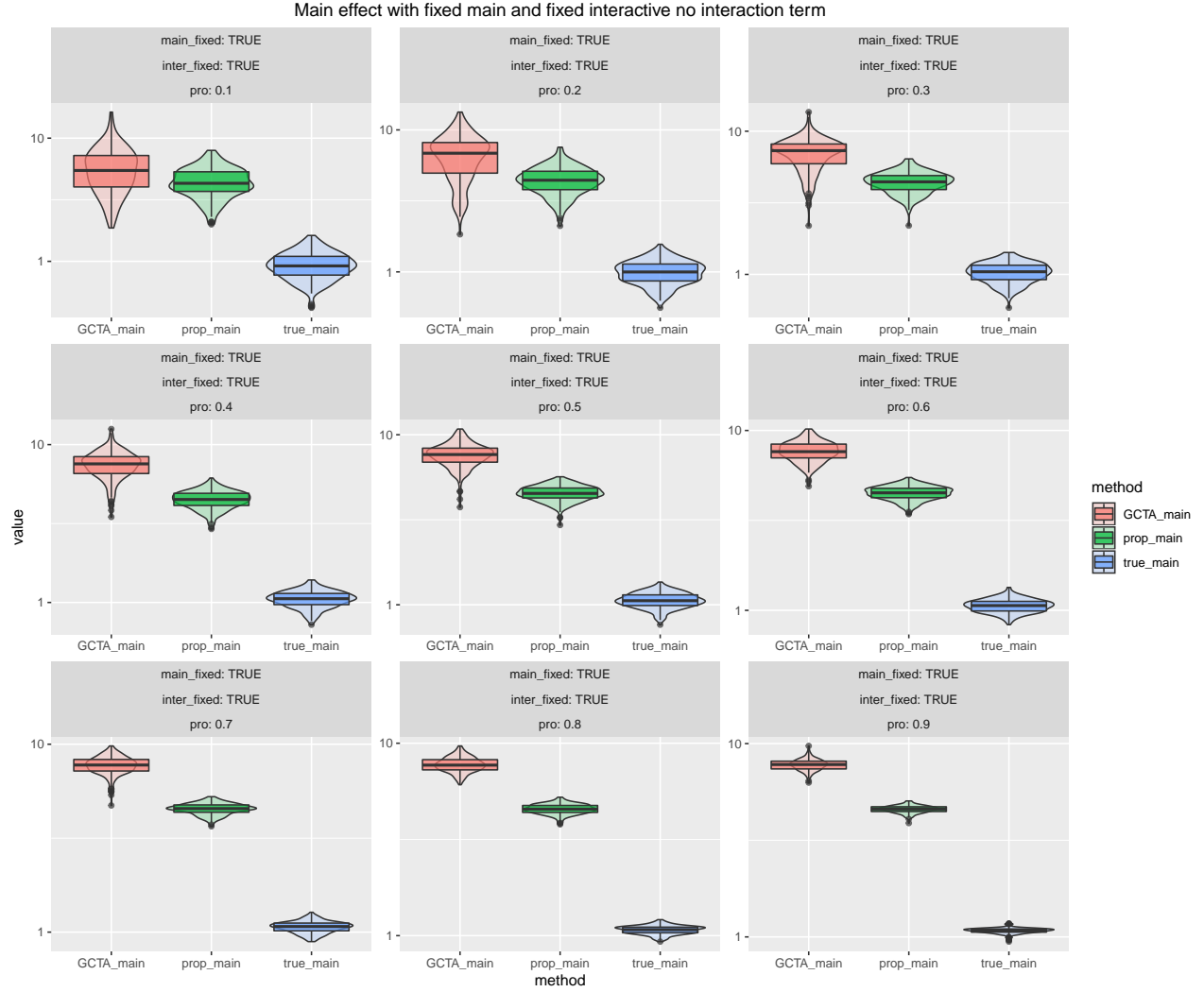

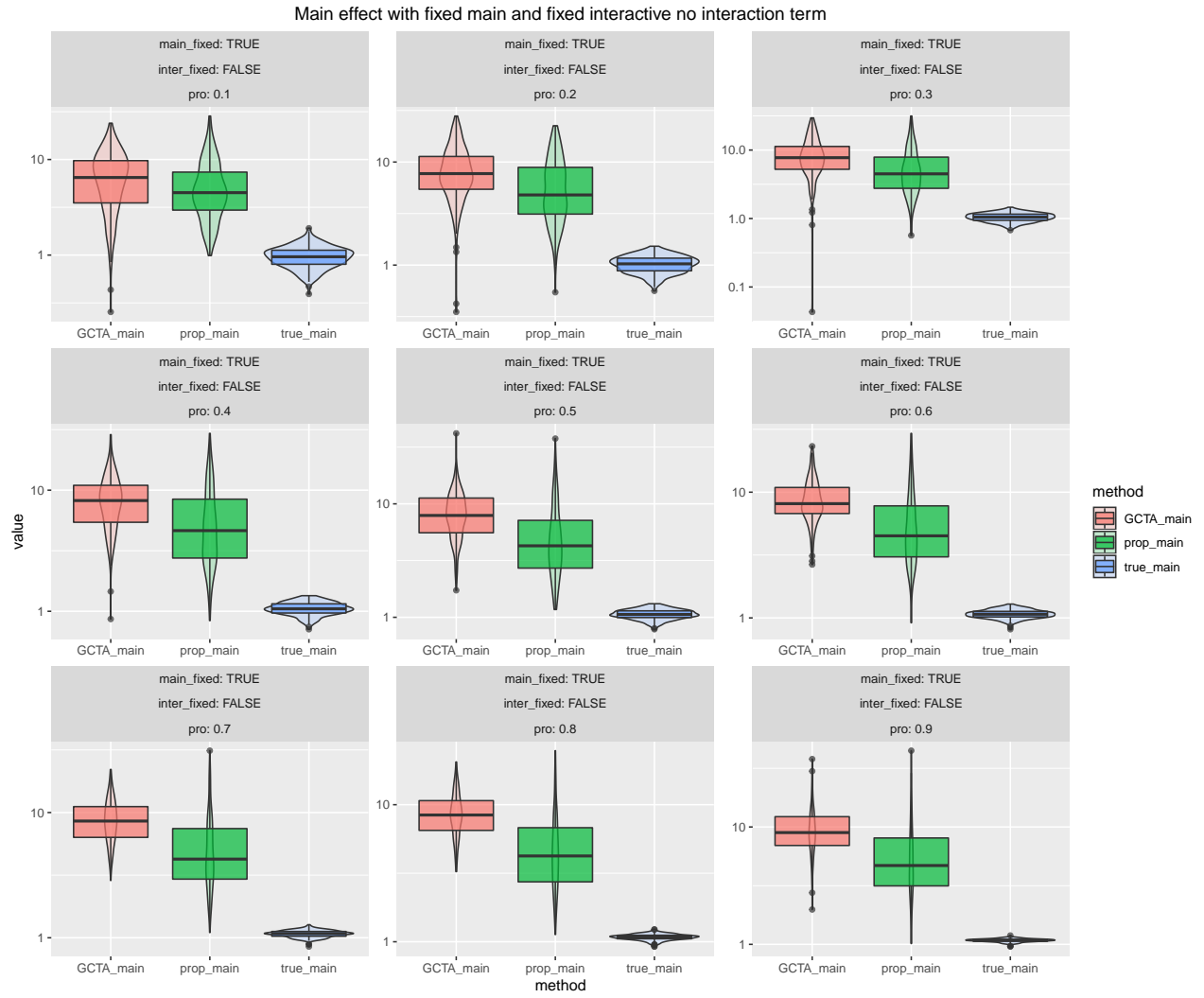Figure 7: Main effect estimation with fixed main and fixed interaction

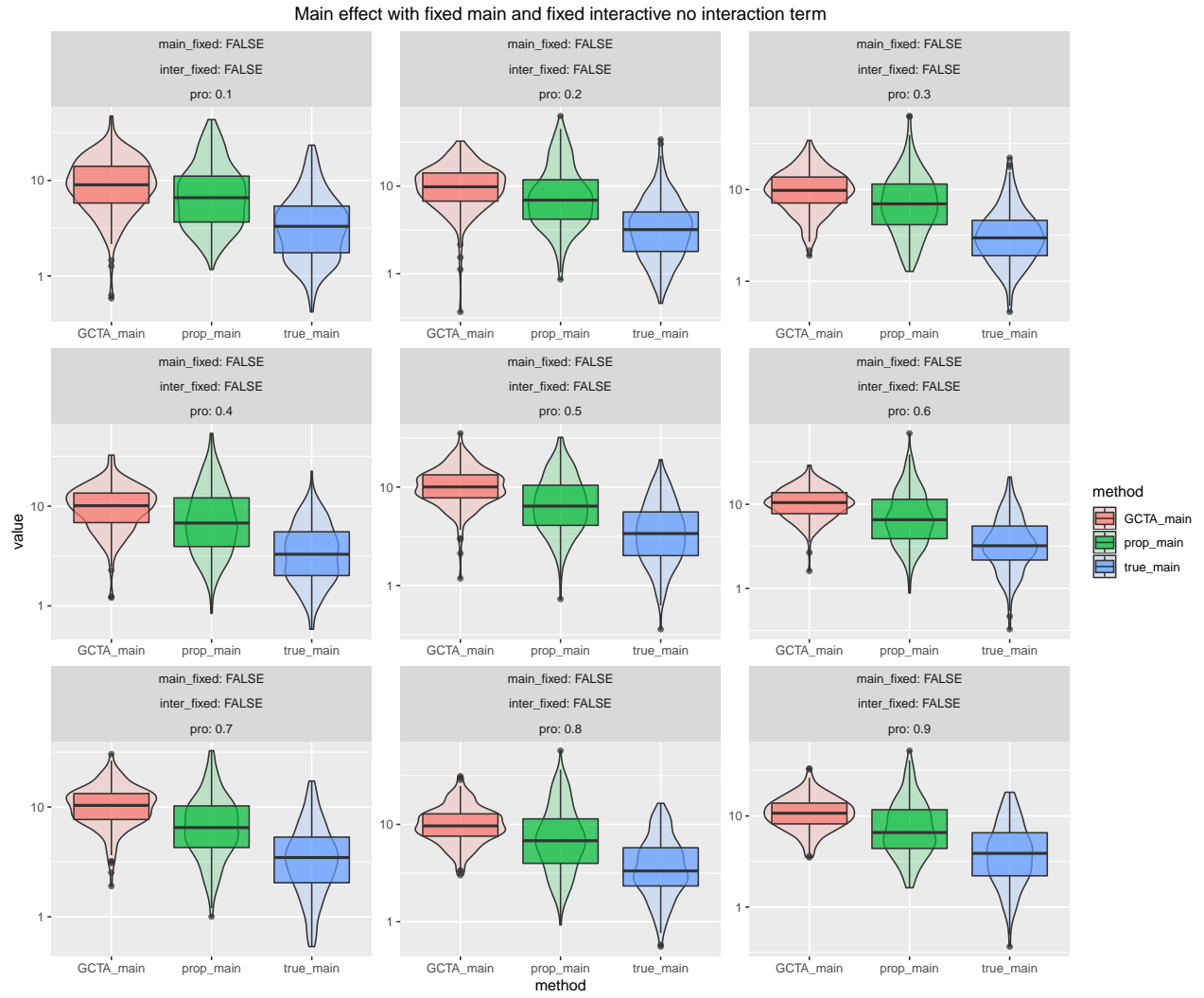Figure 8: Main effect estimation with fixed main and random interaction

Figure 9: Main effect estimation with random main and random interaction

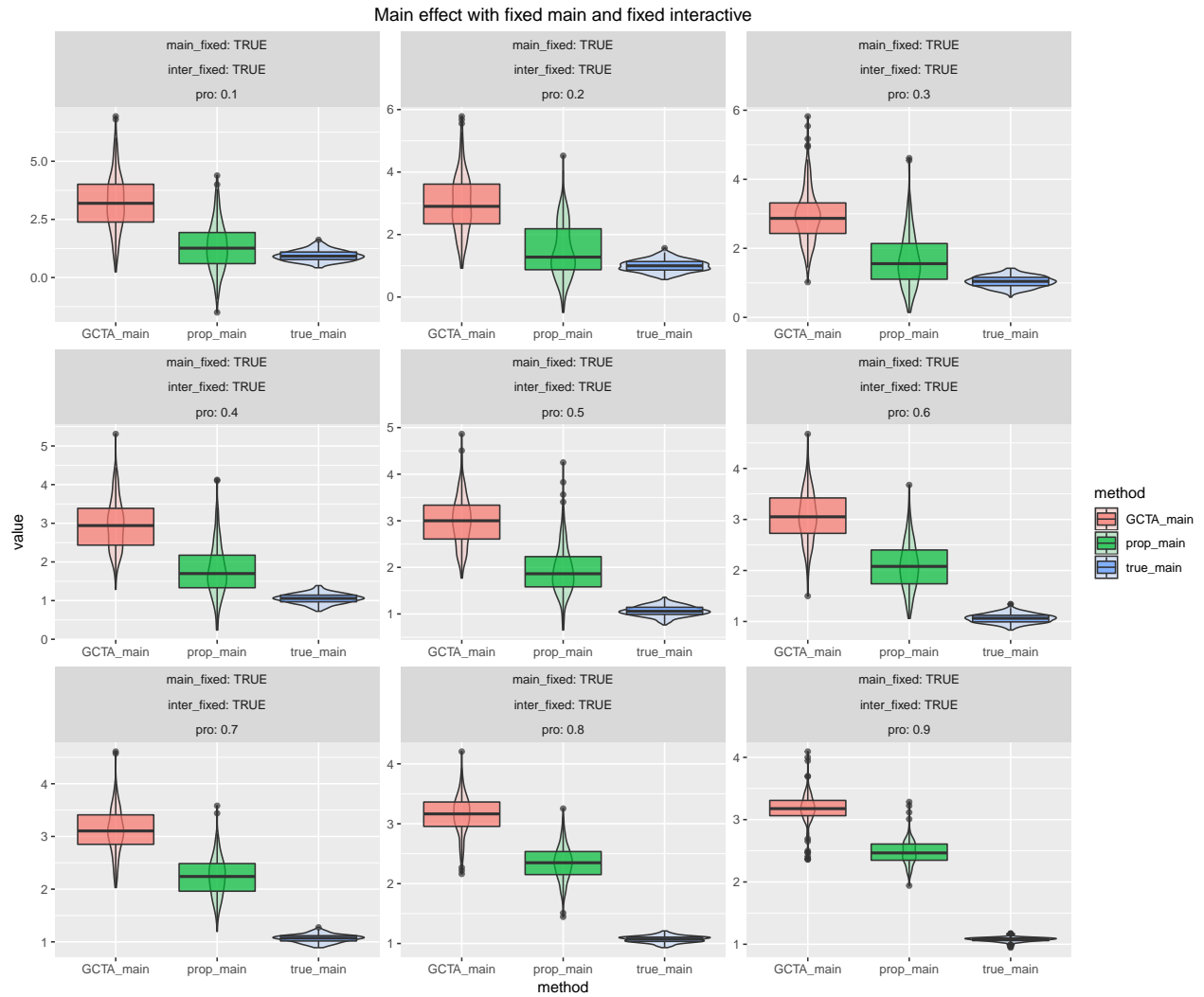## 3.4 Main effect estimation (with interaction term in the model)

Figure 10: Main effect estimation with fixed main and fixed interaction
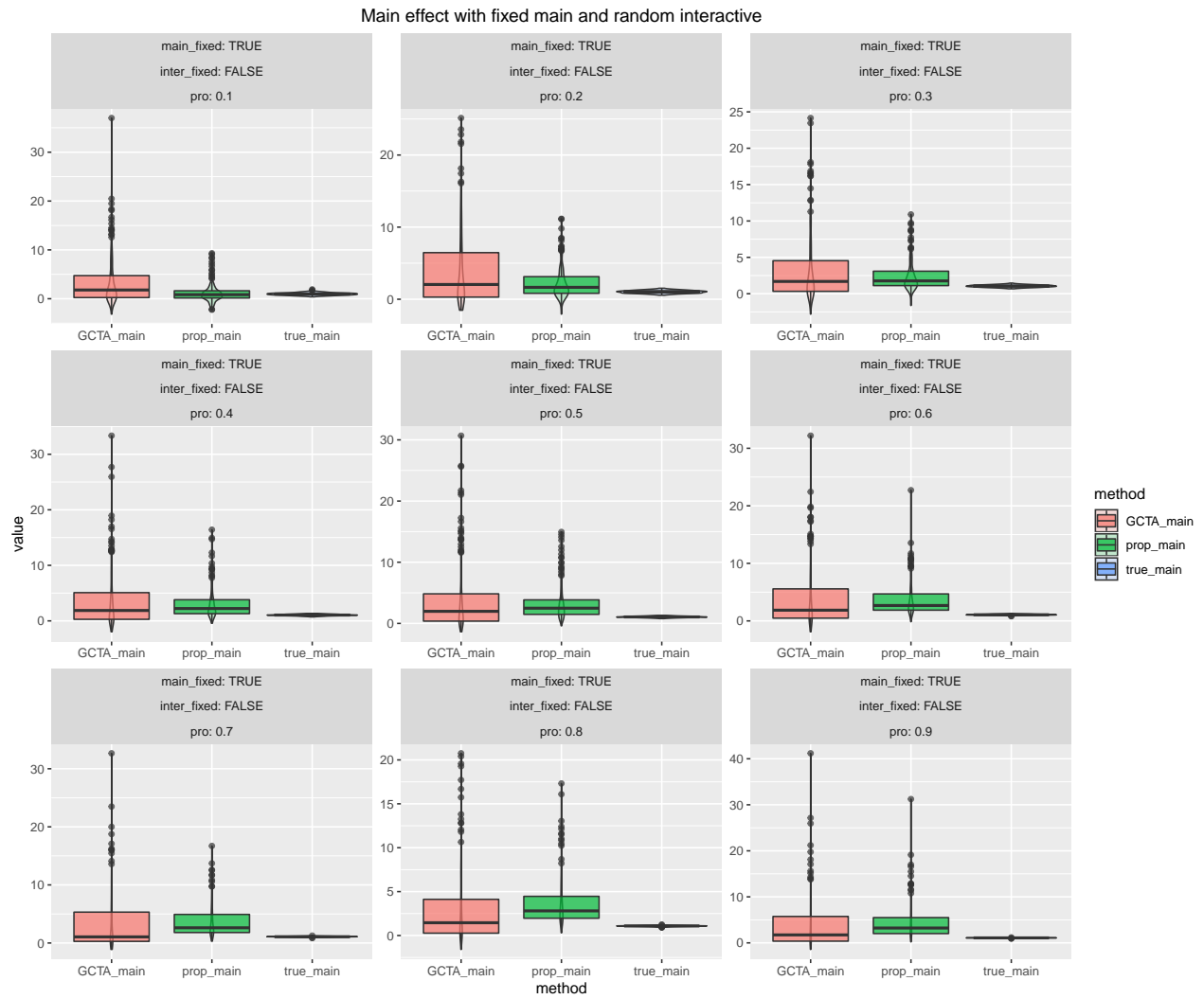
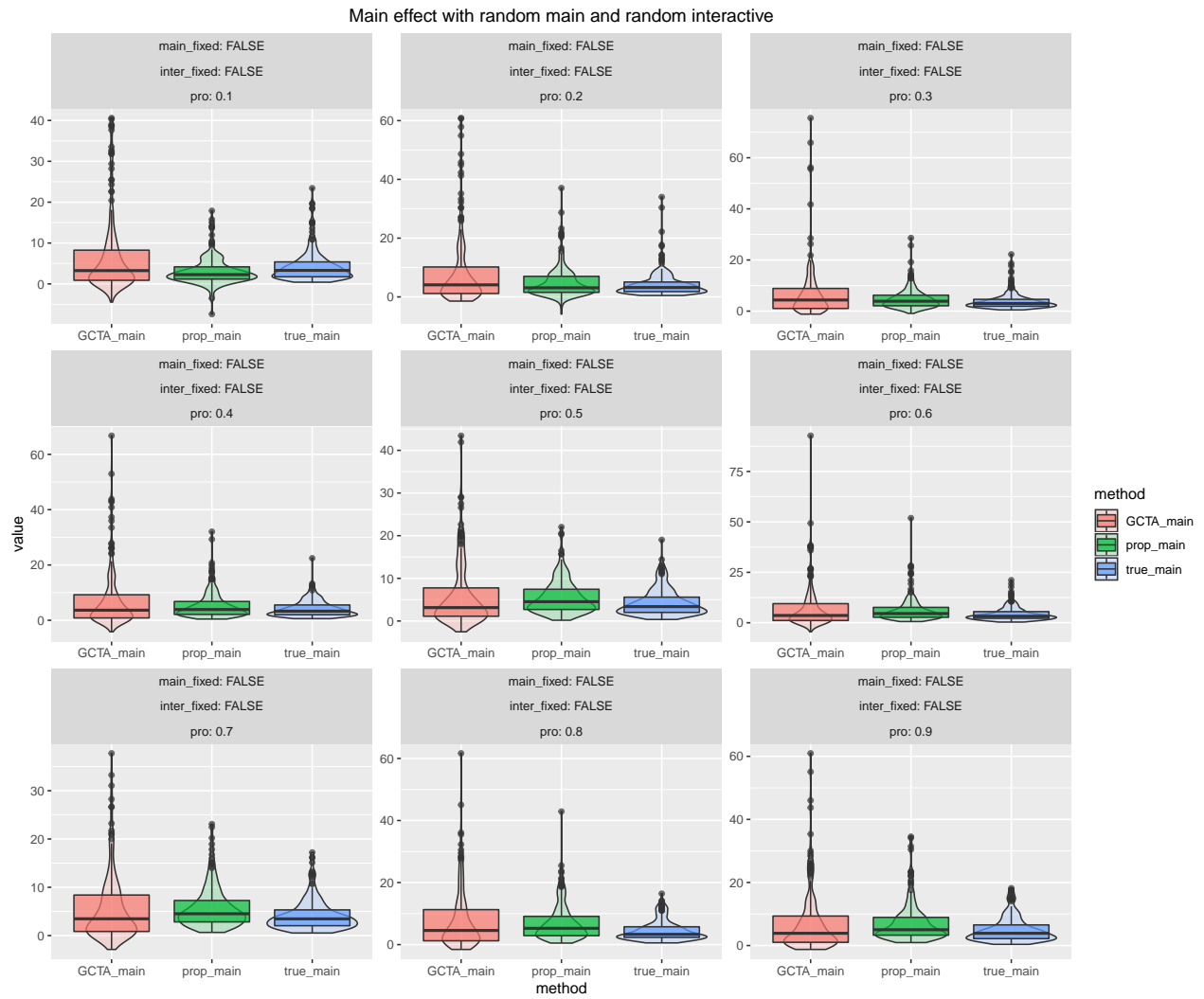Figure 11: Main effect estimation with fixed main and random interaction

Figure 12: Main effect estimation with random main and random interaction

4 Conclusion

5 Further work