Simulation of fixed and random effect

Xuelong Wang

2018-09-27

Contents

1 Motivation

Based on the previous simulations results, we found that the proposed method is able to work well on estimating the interactive effect in certain situation. However, because of the interaction terms the covariates cannot be **Independent** anymore. Thus, we may need to conduct a series simulations with different setup in order to get a large picture of what situation the proposed method can work well.

2 Background

2.1 GCTA method and proposed method

In the GCTA method, the author suggests that the causal covariates should be independent themselves and also be independent with non-causal covariates. Therefore, to ensure the GCTA method is able to work appropriately, we need to make sure the independence among the covariates.

The approach of the proposed method is to uncorrelated the observed covariates by using a linear transformation, i.e. $Z = XA^{-1}$. Given $X \sim N(0, \Sigma)$, we can using the proposed method to get a independent covariates. Thus, this method should work well when X's columns are correlated to each other, which is also supported by simulation results on the main effects with normal correlated distributions.

How to reduce the correlation (the linear transformation) is important part of the proposed method. Because we want to the decorrelated the covariates so that they become independent. We don't want to pick up to much noise during the decorrelation procedure. This is not the main topic of this simulation, but we will do a more details study on the decorrelation.

2.2 Adding the interactive terms

From the model point of view, it actually doesn't change a lot. We can consider the interactive term as another part of covariates. However, since the interactive effects have a strong relation with main effect, that causes the estimation of variance of interactive effect complicated.

2.2.1 Issues

$$\begin{aligned} Var(Y_i) &= Var(X_i^T\beta + X_i^T\Gamma X_i) + Var(\epsilon_i) \\ &= Var(X_i^T\beta) + Var(X_i^T\Gamma X_i) + 2Cov(X_i^T\beta, X_i^T\Gamma X_i) + Var(\epsilon_i) \end{aligned}$$

1. There is an additional terms $Cov(X_i^T\beta, X_i^T\Gamma X_i)$
2. The main effect X_i and X_iX_j are dependent and cannot be independent anymore. Besides, X_iX_j and $X_iX_{j'}$ are dependent.
3. In order to keep the variance structure, we can only apply the linear transformation on the main effects, not the interactive effects.

2.2.2 Solutions

For the 1st issue, there are basic two different approach.

1. If X or after some transformation follows normal distributions, then the $Cov(Z_i^T\beta, Z_i^T\Gamma Z_i)$ will be zero. More specifically, as long as the 3rd moments are 0 the correlation of main and interactive effect should be zero.
2. If we assume $\beta's$ or $\Gamma's$ are independent random variables, then the correlation should also be round zero. Because we assume $\beta \perp\!\!\!\perp \Gamma$

For the 2nd issue, the simulation studies shows that the the proposed method can work well under the normal distribution setup. So it suggests that the Independence assumption may not be that restricted, at least under the Normal distribution. One guess is that under normal distribution the covariates of main and interactive is zero $Cov(X_i^T\beta, X_i^T\Gamma X_i) = 0$, which means they are un-correlated. Although un-correlation can not guarantee independence, it shows non-linear relation between main and interactive effects.

Thus, the main idea is to first transform the covariates into normal-like distribution and applied the proposed method.

3 Simulation

To get a better and larger picture of how well the proposed method's estimations, we conduct a series simulation with different setup.

The main factors are

1. Distribution of X

2. Independence or Dependence of X
3. Fixed or Random of main effect $\beta' s$
4. Fixed or Random of interaction effect $\Gamma' s$

3.1 Simulation procedure

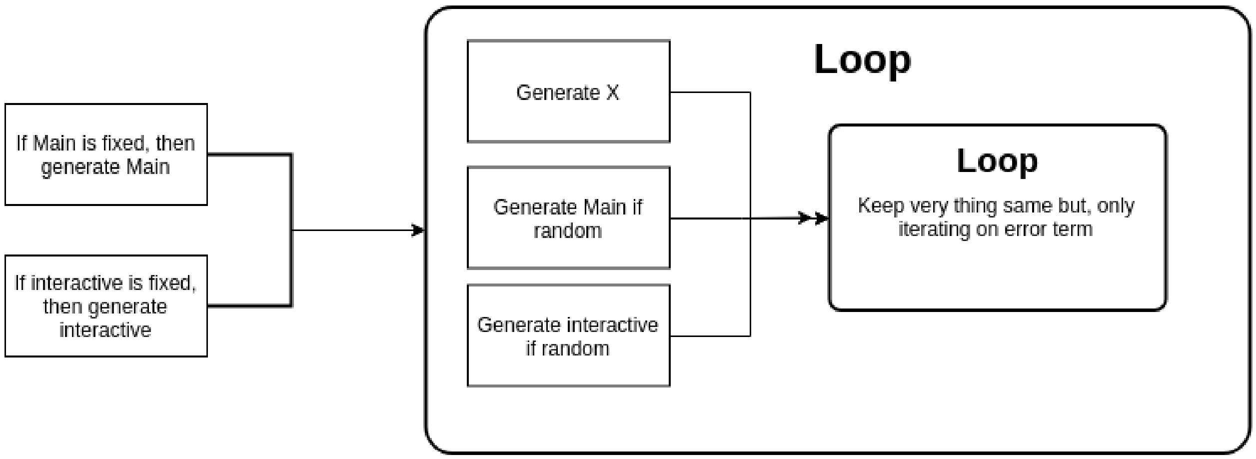

Figure 1: Simulation workflow

4 X follows Normal distributions

If we can assume that X has normal distribution then the problem become more straightforward, because the correlation of main and interactive effects are zero. Besides, we don't need to transform X in order to make it looks like a normal distribution.

4.1 Independent

Based on the previous simulation (from the Proposal), if X is normal and independent, then both proposed and original GCTA method can estimate the main and interaction effects well. This is indirectly suggest that the dependence between main and interactive may not be a big issue.

4.2 Dependent (correlated)

If the X's are normal but correlated with each other, then it's not easy to estimate the interactive correctly. Based on previous simulation's results, the proposed method can work much better than the original GCTA when β, the main effect, is fixed. However when the main coefficients are not fixed then there is a big bias for the interaction estimation of proposed method. A guess is that if the β is not linear transformation of may affect the $Var(X_i^T \Gamma X_i)$.

4.2.1 β is fixed and Γ is fixed

The estimation and true values

$$Var(X_i^T\beta) = \beta^T\Sigma\beta$$

$$Var(X_i^T\Gamma X_i) = 2tr(\Gamma\Sigma\Gamma\Sigma) + 0$$

Simulation result

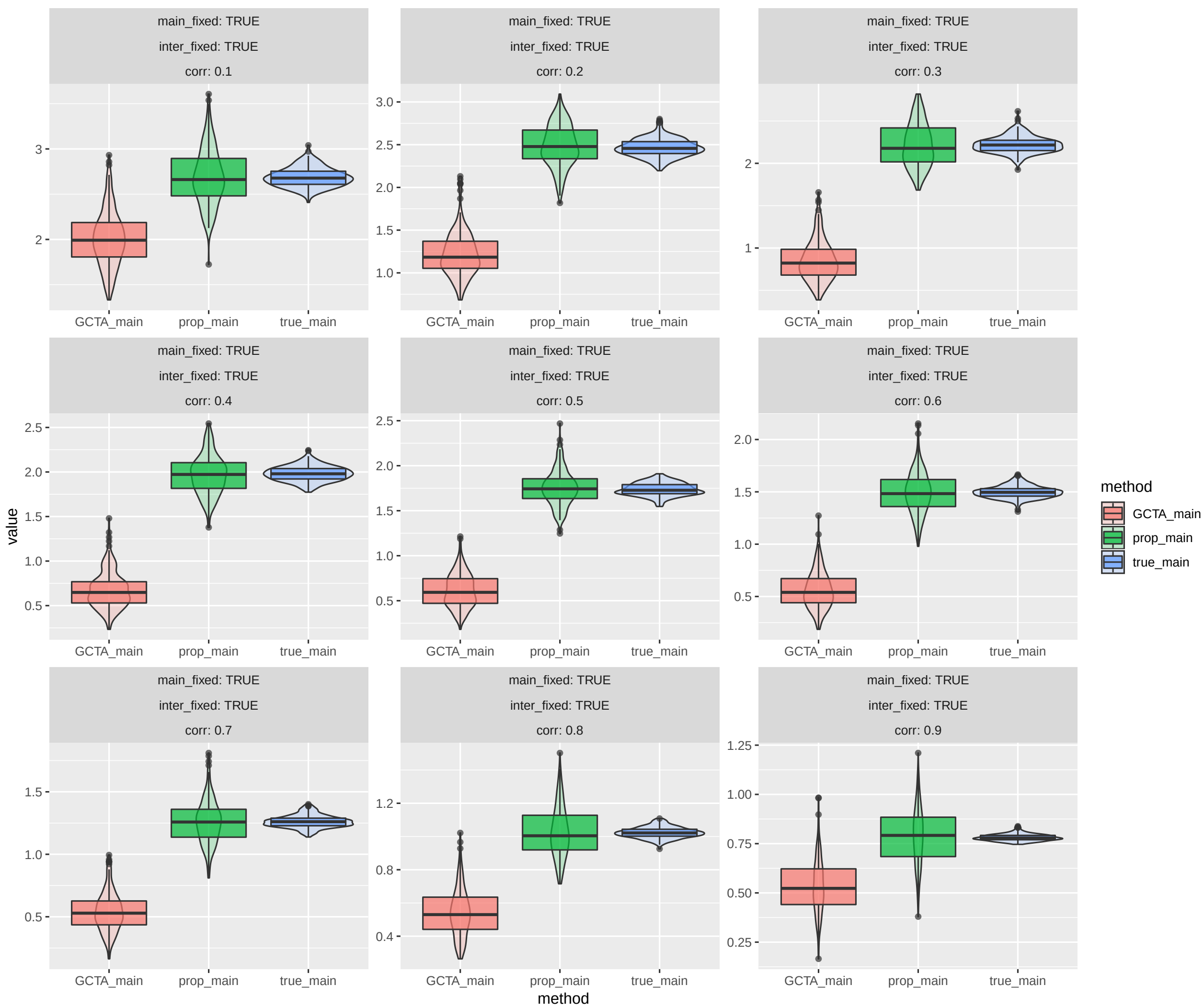

Figure 2: Main Effect of normal fixed main and interaction

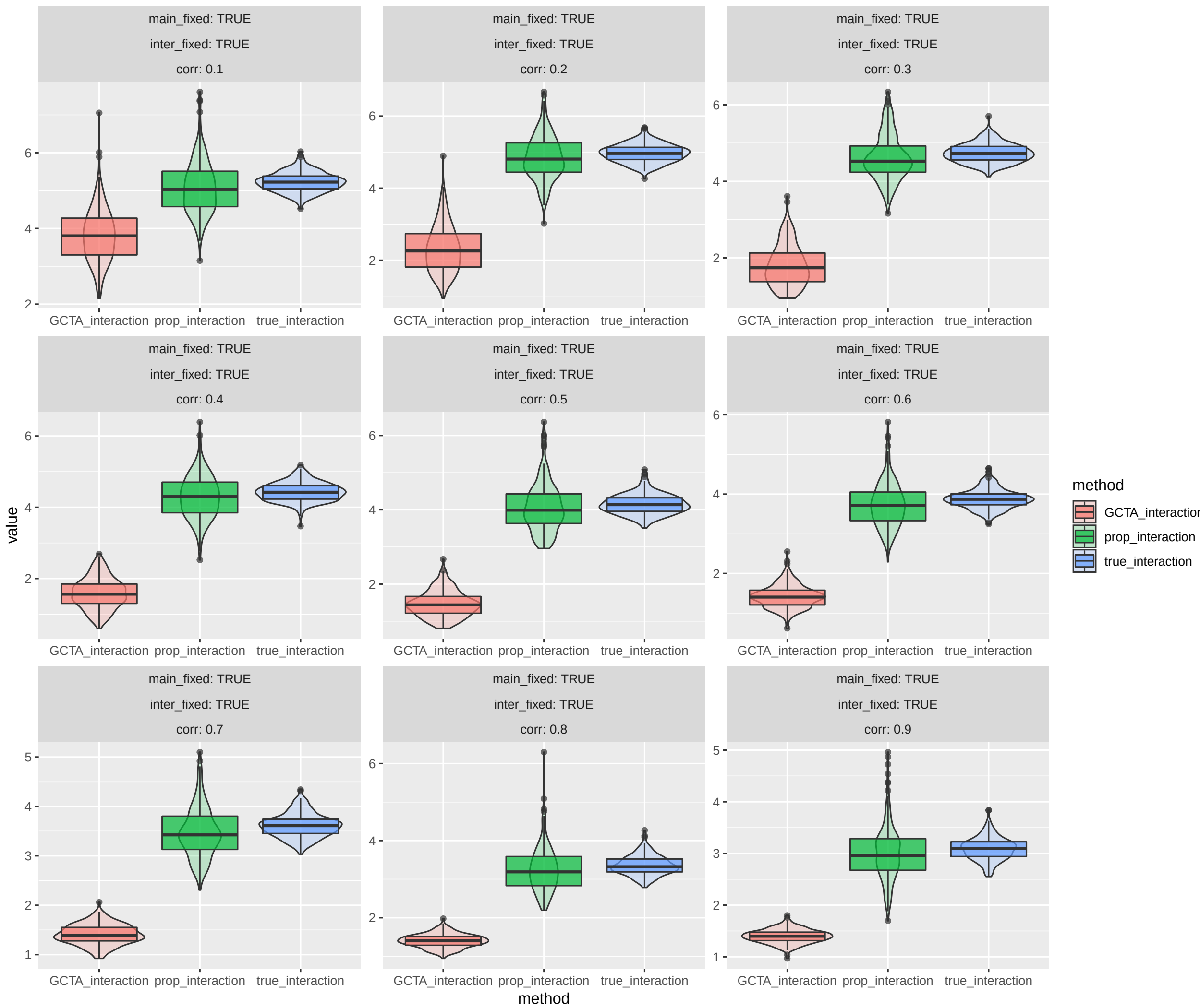

Figure 3: Interactive Effect of normal fixed main and interaction

4.2.2 β is fixed and Γ is random

$$Var(X_i^T\beta) = \beta^T\Sigma\beta$$

$$\begin{aligned} Var(X_i^T\Gamma X_i) &= E(Var(X_i^T\Gamma X_i|X_i = x_i)) + Var(E(X_i^T\Gamma X_i|X_i = x_i)) \\ &= E(Var(X_i^T\Gamma X_i|X_i = x_i)) \end{aligned}$$

Simulation result

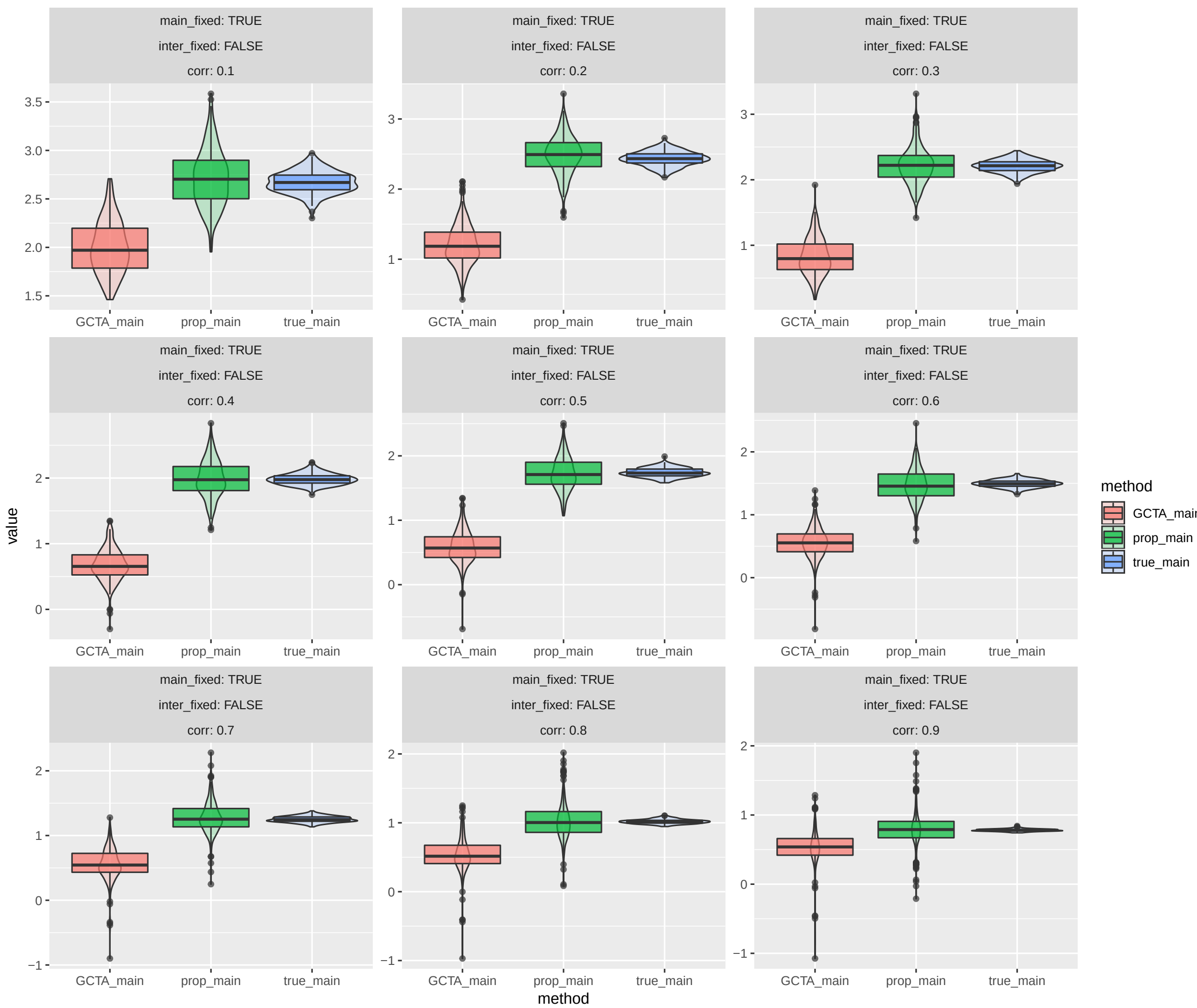

Figure 4: Main Effect of normal fixed main and random interaction

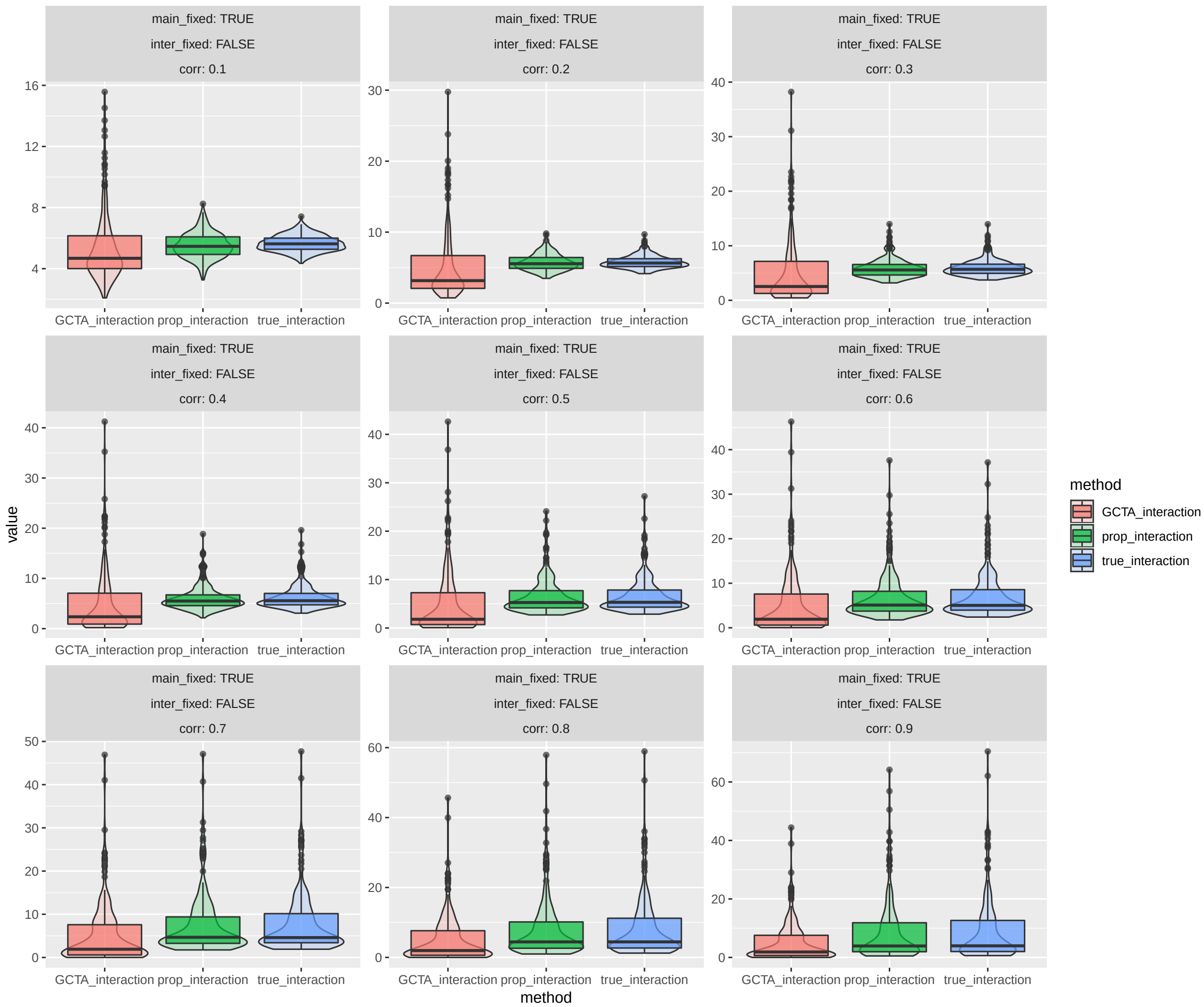

Figure 5: Interactive Effect of normal fixed main and random interaction

4.2.3 β is random and Γ is random

$$
\begin{aligned}
Var(X_i^T\beta) &= E(Var(X_i^T\beta|X_i = x_i)) + Var(E(X_i^T\beta|X_i = x_i)) \\
&= E(Var(X_i^T\beta|X_i = x_i)) + 0 \\
&= E(X_i^T\Sigma_\beta X_i) && \text{If } \Sigma_\beta = I_p \\
&= E(X_i^T X_i)
\end{aligned}
$$

$$
\begin{aligned}
Var(X_i^T\Gamma X_i) &= E(Var(X_i^T\Gamma X_i|X_i = x_i)) + Var(E(X_i^T\Gamma X_i|X_i = x_i)) \\
&= E(Var(X_i^T\Gamma X_i|X_i = x_i))
\end{aligned}
$$

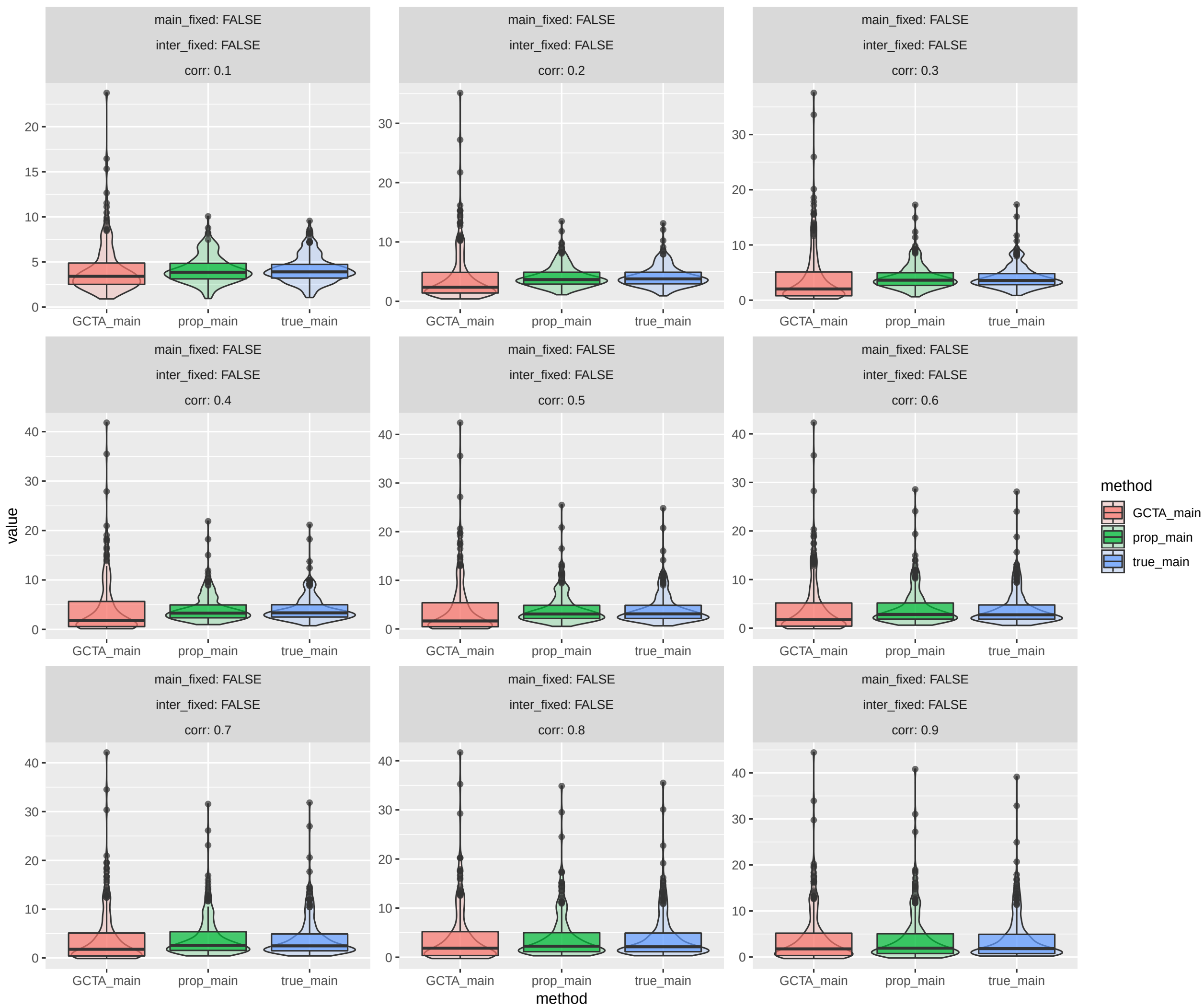

Figure 6: Main Effect of normal random main and interaction

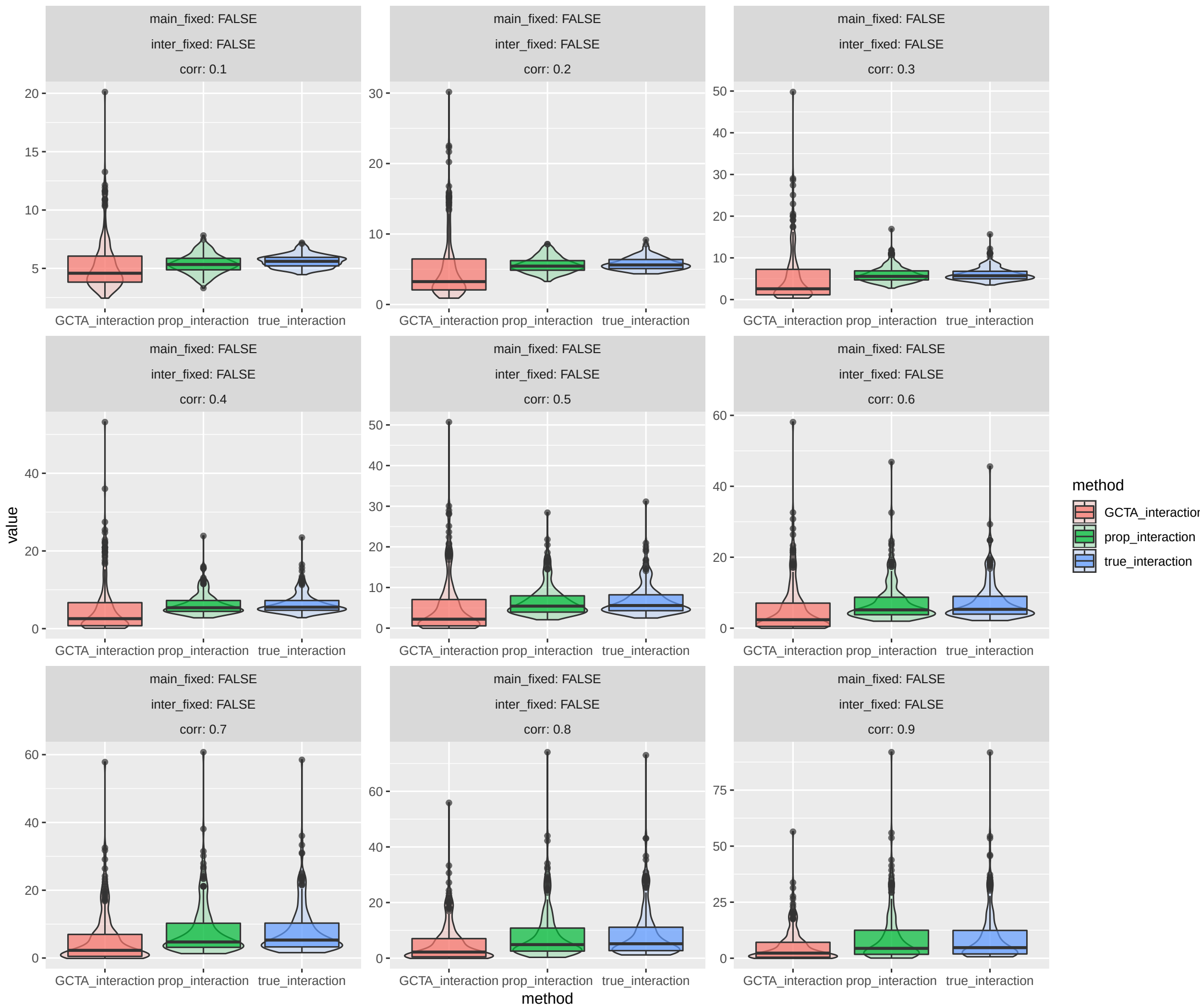

Figure 7: Interactive Effect of normal random main and interaction

5 X follows Chi-square distribution

In this case, we choose a long-tail distribution. Since the 3rd moment of Chi-square is no longer 0, so the $Cov(X_i^T\beta, X_i^T\Gamma X_i)$ cannot be zero also. The following simulation is trying to study how this non-zero term affect the performance of the proposed and original GCTA method.

5.1 Independent

5.2 Dependent

The procedure of the data generation is following:

1. Generate a vector dependent random variable $Z \sim N(0, \Sigma_{p*})$
2. Square each elements of Z, $Z^2 = (Z_1^2, \ldots, Z_{p*}^2)^T$
3. Divid $p*$ into p group and let $X_i = \sum_{k\in i} Z_k^2$, where i all elements in the Ith group.
4. Let $X = (X_1^2, \ldots, X_p^2)^T$

5.2.1 β is fixed and Γ is fixed

Simulation result

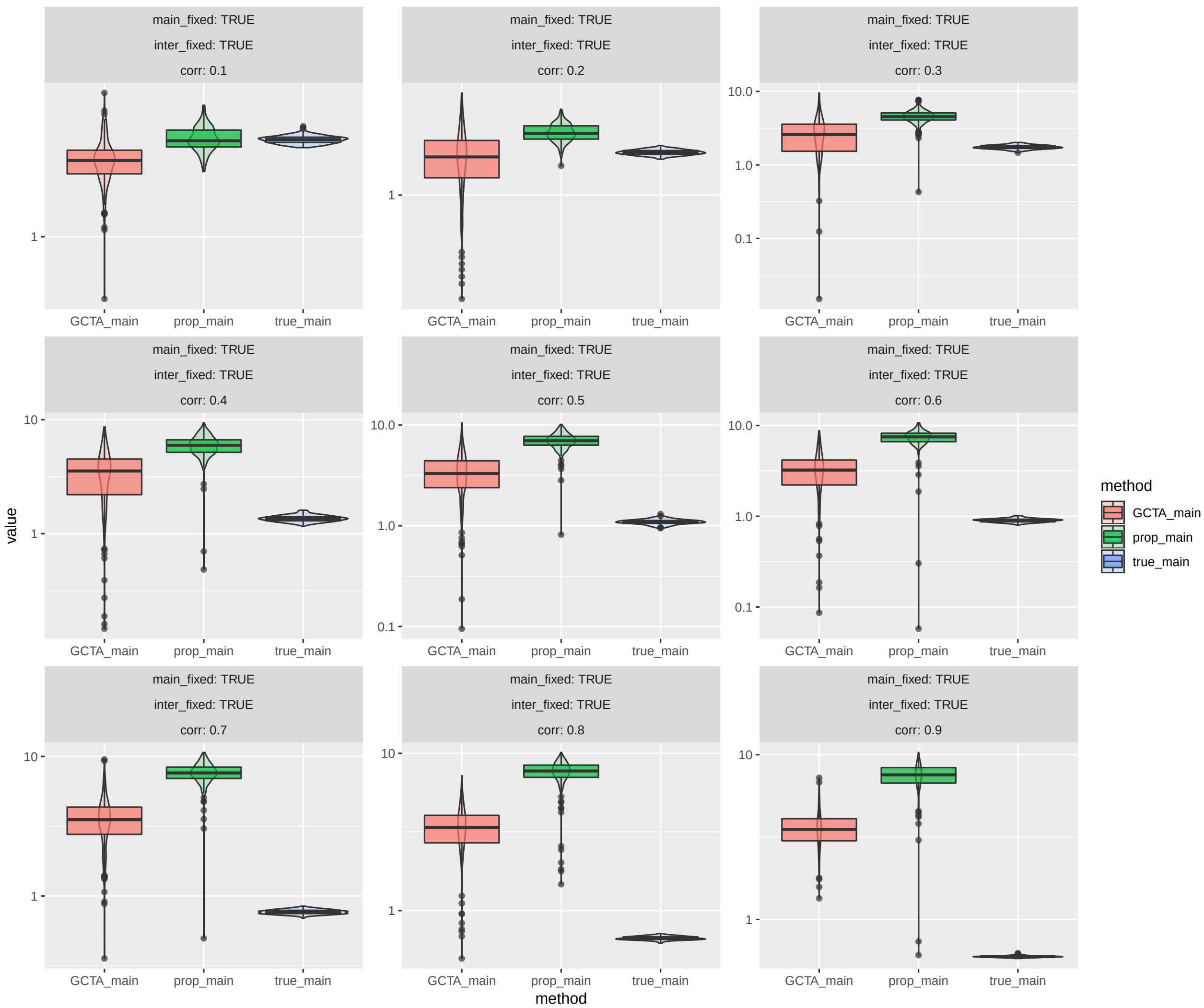

Figure 8: Main Effect of chi fixed main and interaction

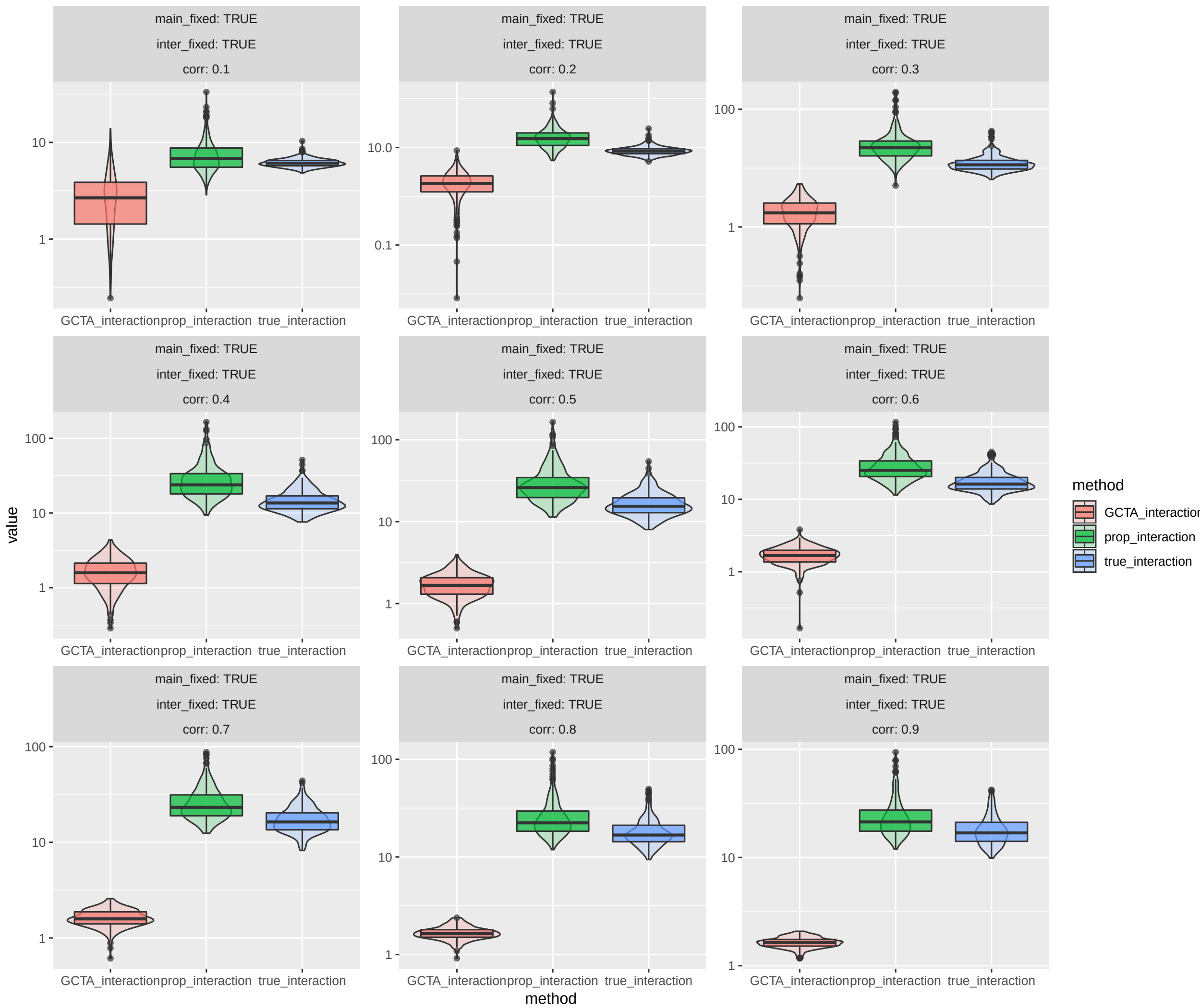

Figure 9: Interactive Effect of chi fixed main and interaction

5.2.2 β is fixed and Γ is random

Simulation result

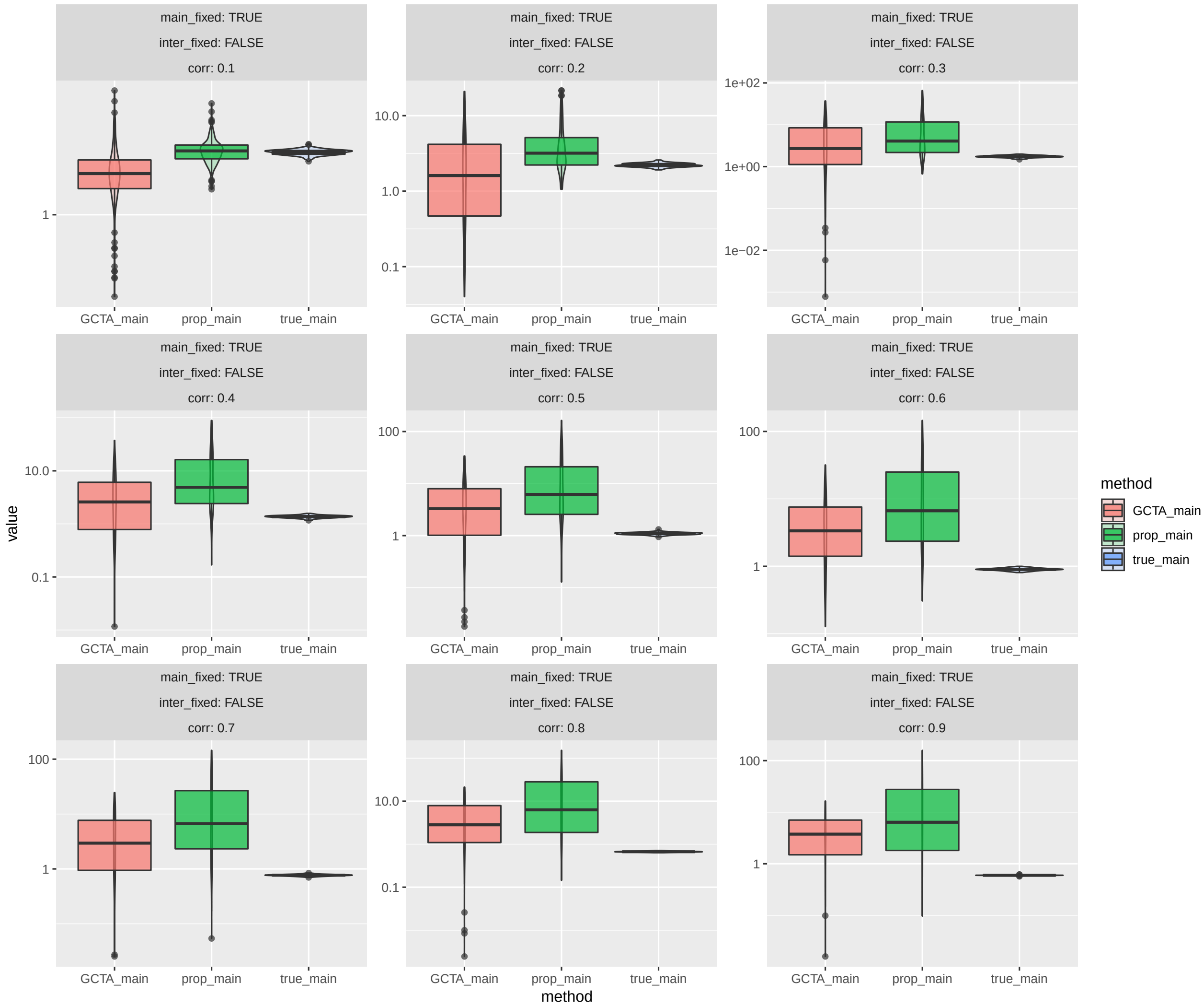

Figure 10: Main Effect of chi fixed main and random interaction

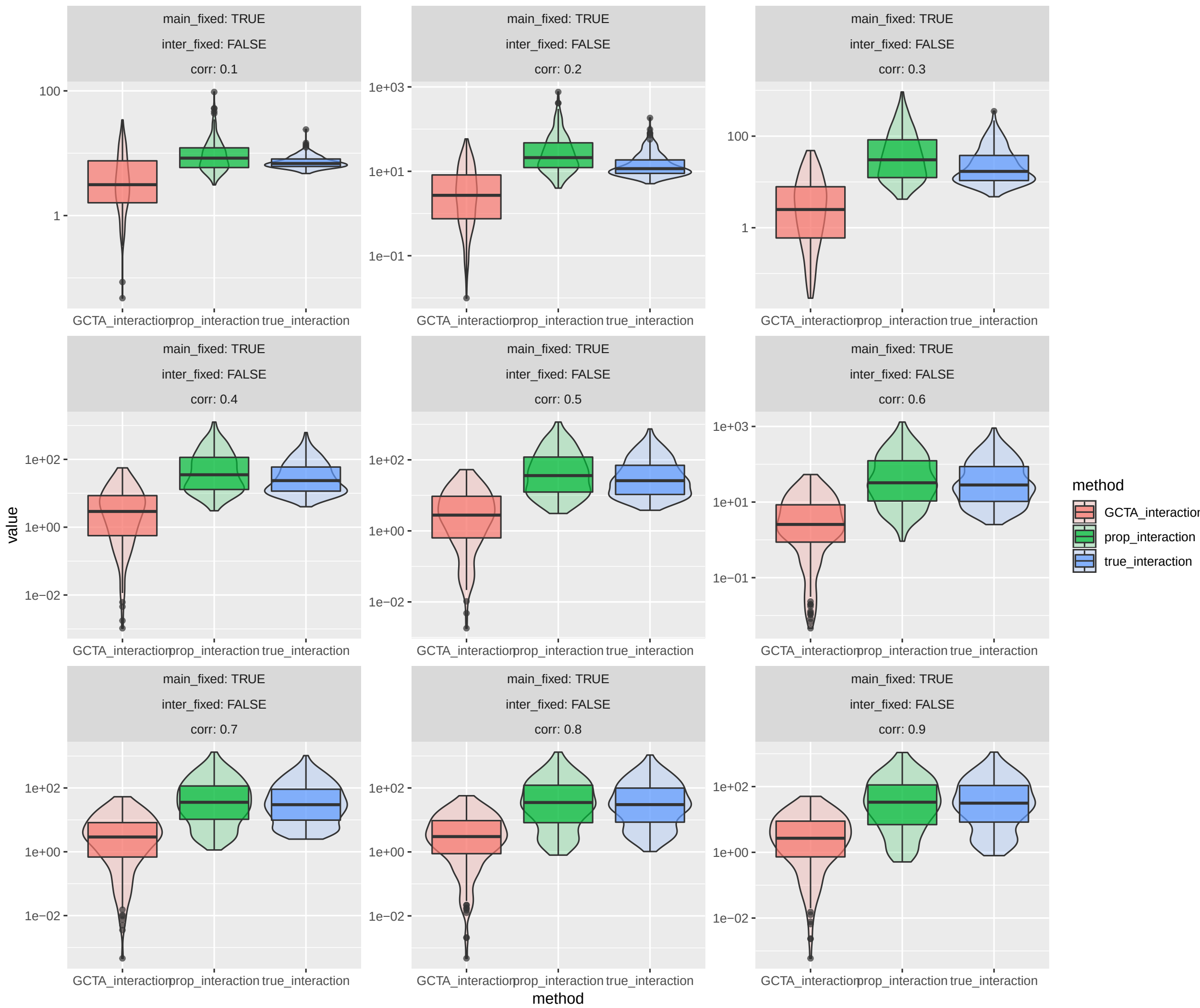

Figure 11: Interactive Effect of chi fixed main and random interaction

5.2.3 β is random and Γ is random

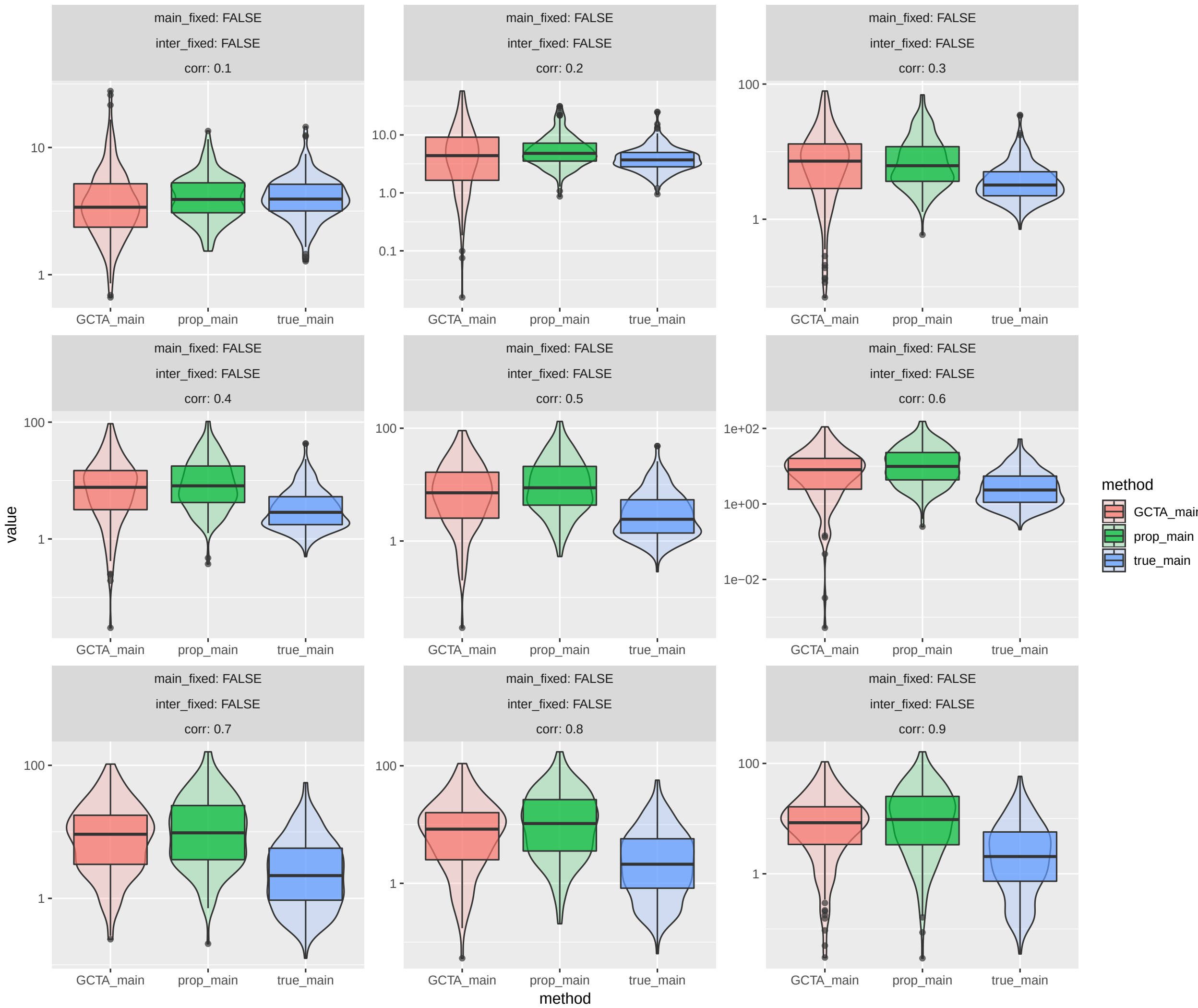

Figure 12: Main Effect of chi random main and interaction

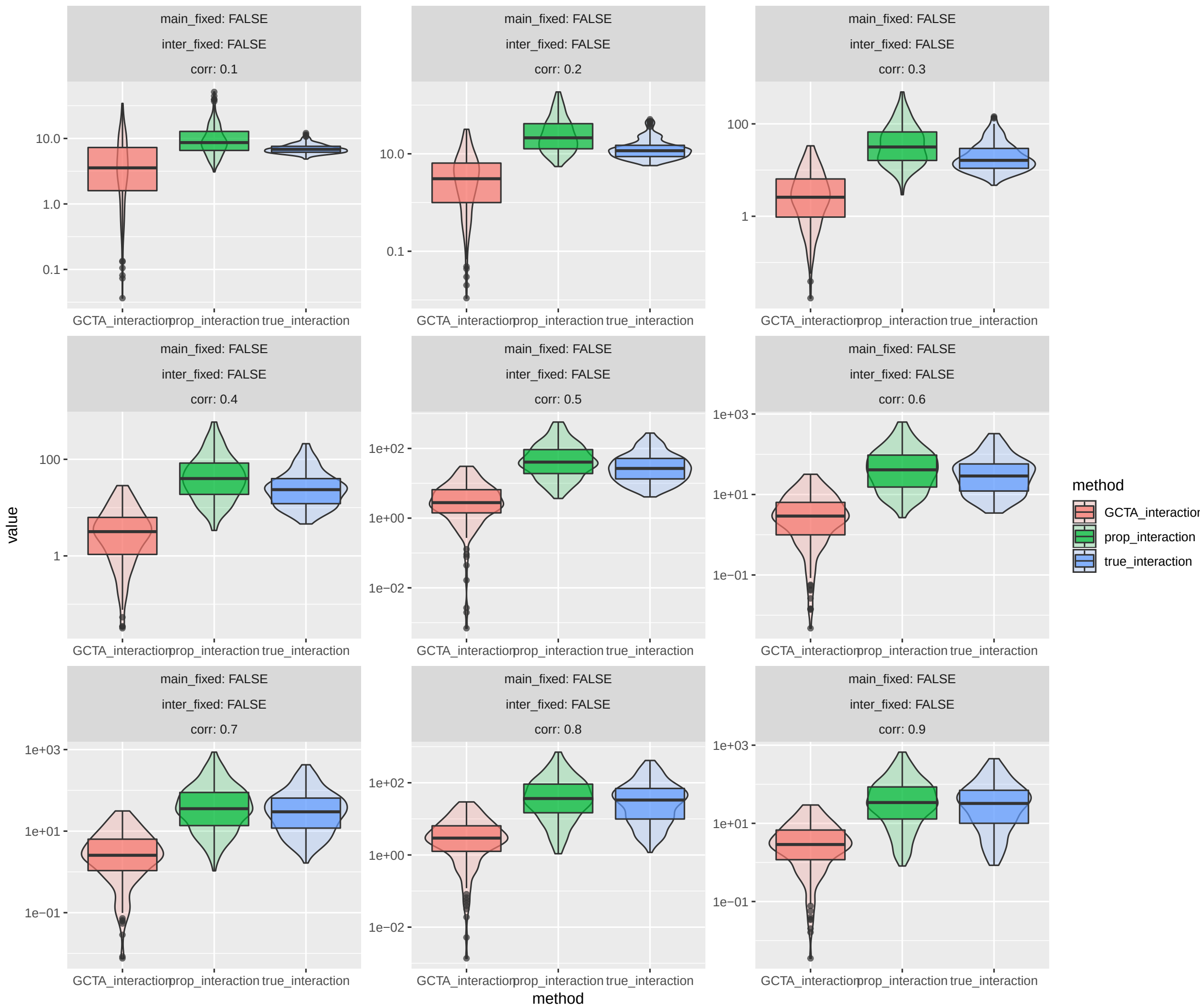

Figure 13: Interactive Effect of chi random main and interaction

5.3 Combine the main and interaction effect

Based on the simulation result of the Chi-square distribution, the estimation of main and effect are not bad, which means the bias of the estimations are not large. However, there does exist some bias because of the non-zero covariance term ($Cov(X_i^T\beta, X_i^T\Gamma X_i)$).

One possible solustion or walk-around method is to estimate the total variance of signal instead of main and interactive variance separately. Therefore, we don't have to worry about the covariance term between the main and interative terms.

More specifically, we treat main and interactive covariates as a long vector and try to estimate its corresponding variance.

$$
\begin{aligned}
Var(Y_i) &= Var(X_i^T\beta + X_i^T\Gamma X_i) + Var(\epsilon_i) \\
&= Var(X_i^T\beta + X_{i,inter}^T\gamma) + Var(\epsilon_i) \\
&= \begin{bmatrix} X_i \\ X_{i,inter} \end{bmatrix}^T \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + Var(\epsilon_i) \\
&= Var(X_{i,total}^T\beta_{total}) + Var(\epsilon_i)
\end{aligned}
$$

The big difference of this approch is that the interaction terms are also standardized and decorrelated.

Do we still have the following equation hold?

$$
Var(X_{i,total}^T\beta_{total}) = Var(Z_{i,total}^T\alpha_{total})
$$

5.3.1 β is fixed and Γ is fixed (Chi df = 10)

Simulation result

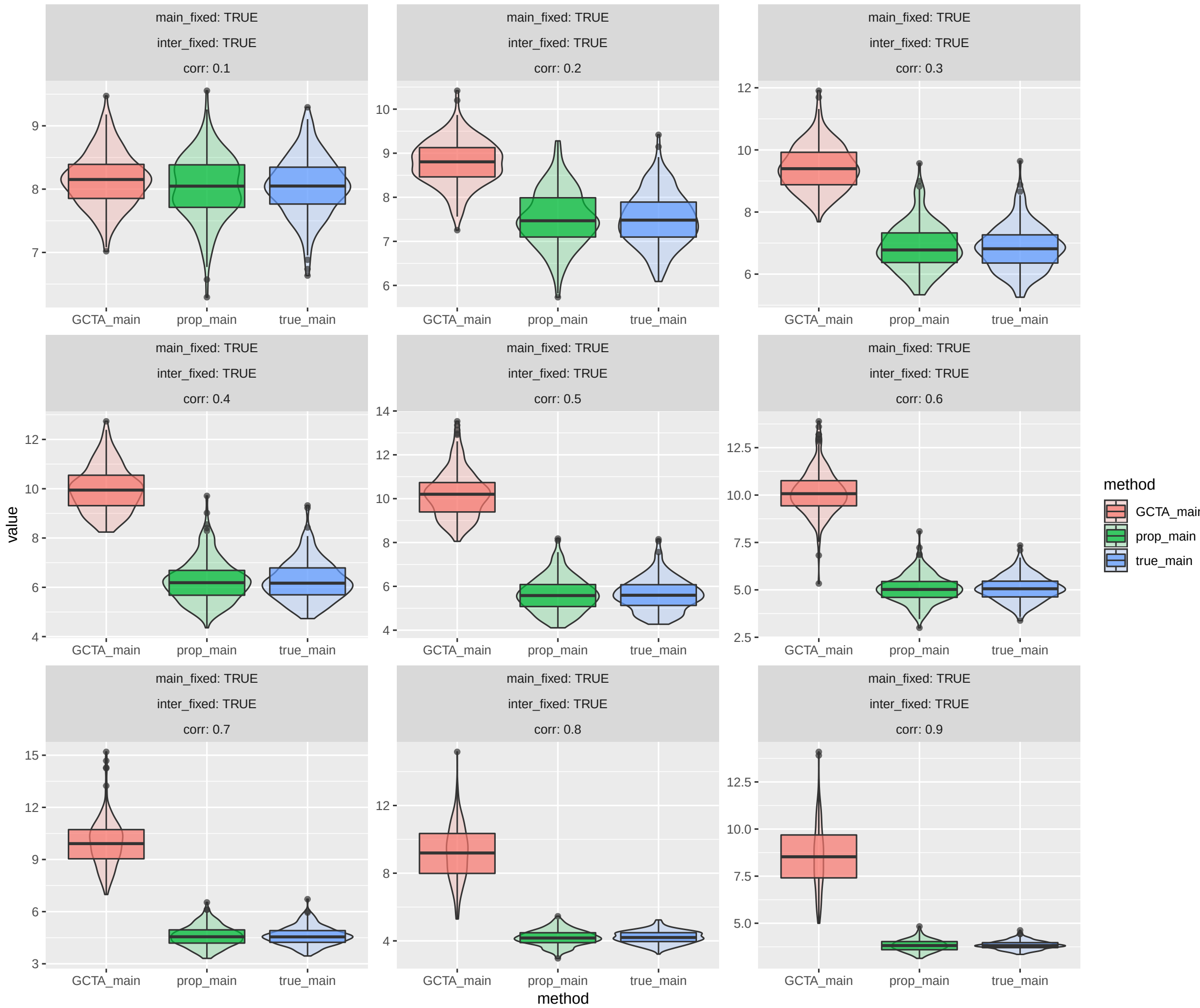

Figure 14: Total effect of chi fixed main and interaction

5.3.2 β is fixed and Γ is random (Chi df = 10)

Simulation result

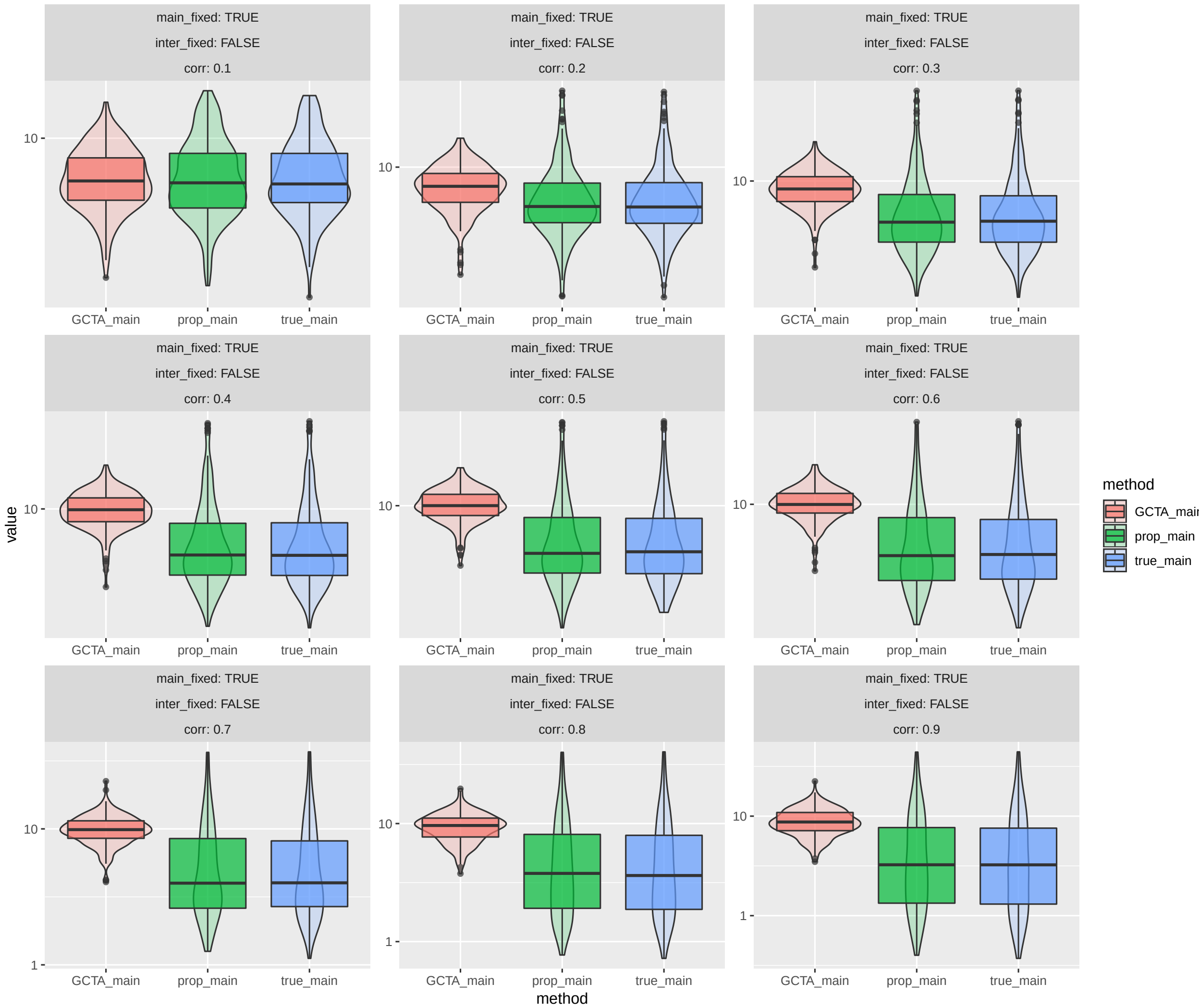

Figure 15: Total effect of chi fixed main and random interaction

5.3.3 β is random and Γ is random (Chi df = 10)

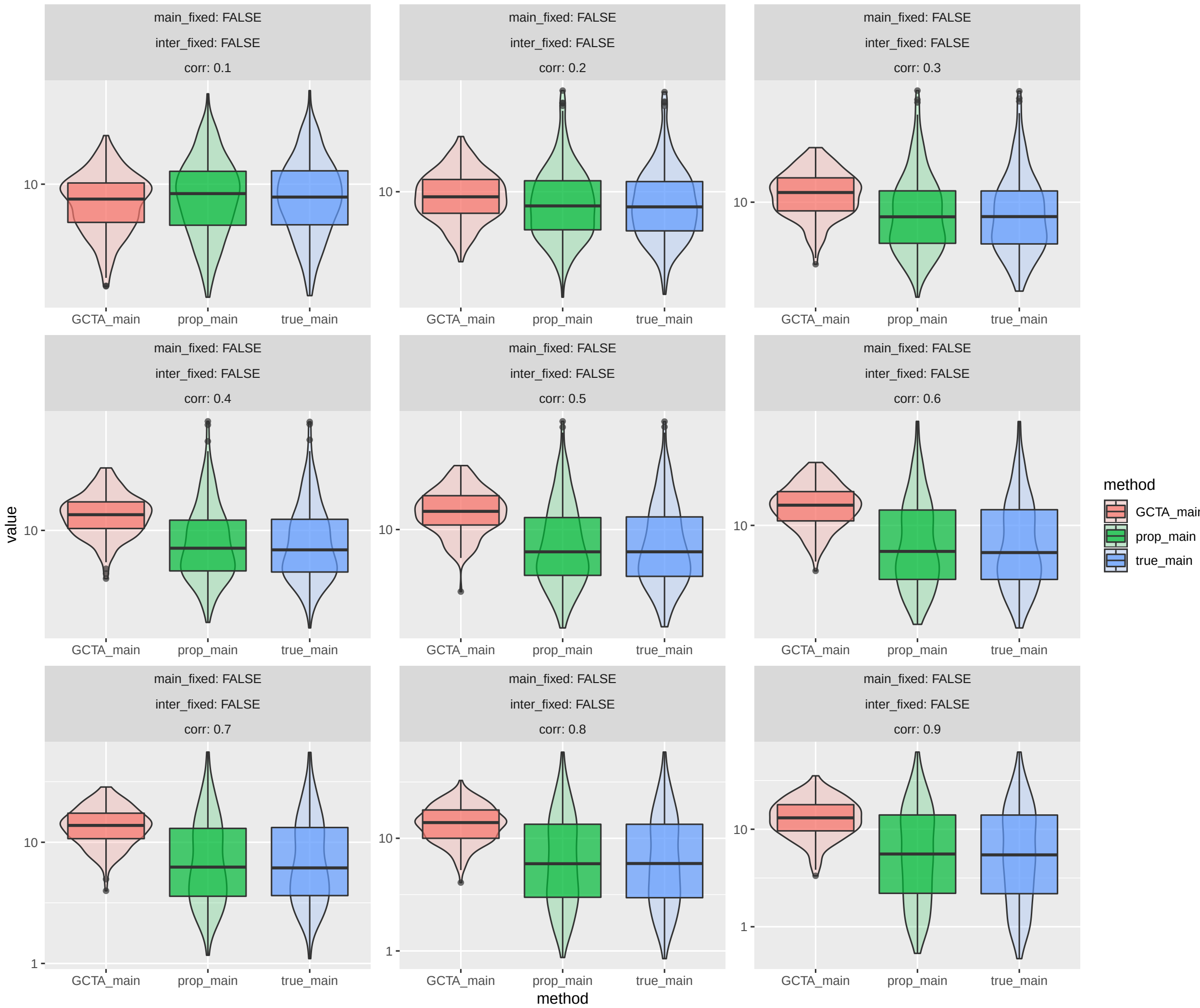

Figure 16: Total effect of chi random main and interaction

5.4 Combine the main and interaction effect with df = 1

5.4.1 β is fixed and Γ is fixed (Chi df = 1)

Simulation result

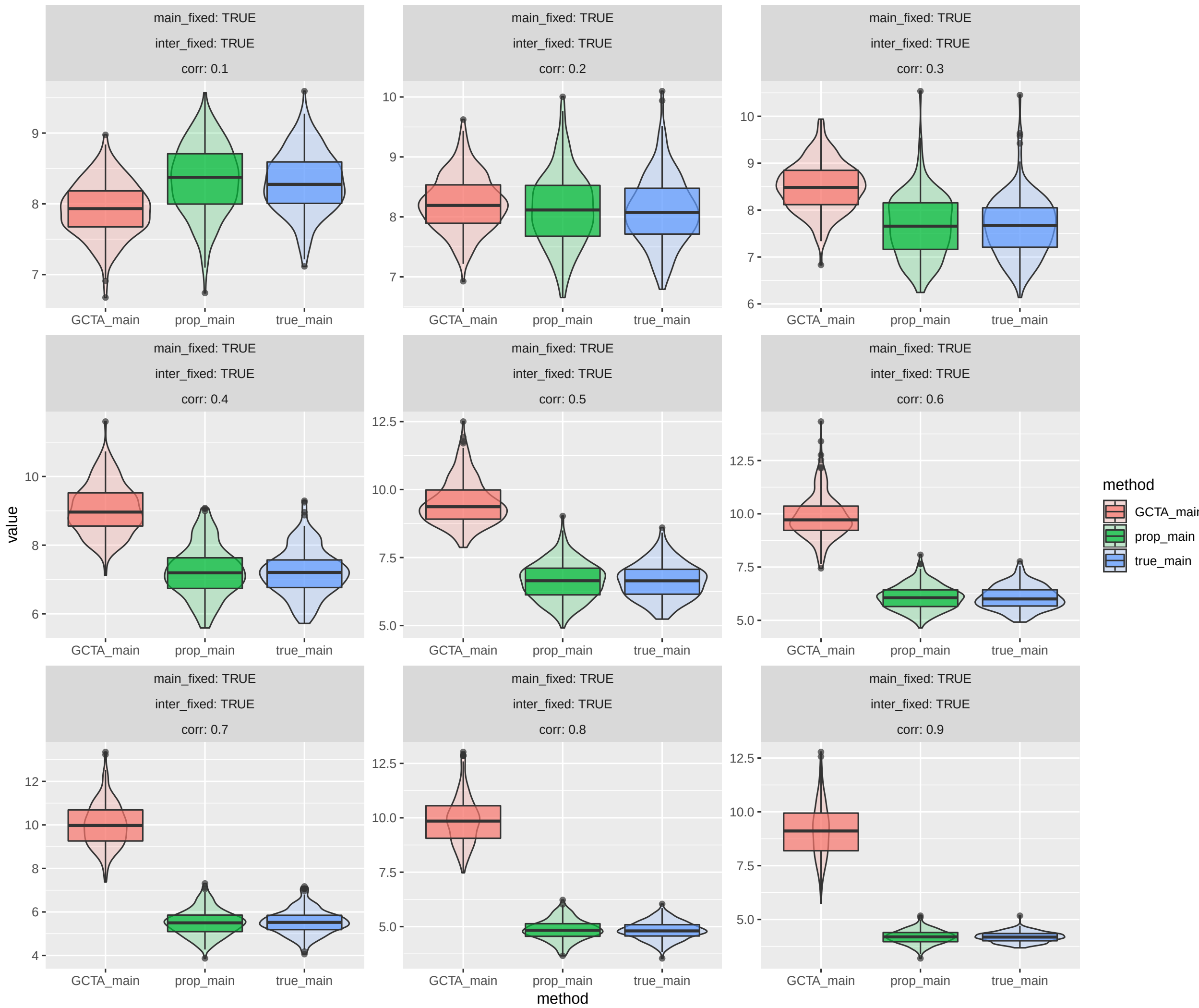

Figure 17: Total effect of chi fixed main and interaction

5.4.2 β is fixed and Γ is random (Chi df = 1)

Simulation result

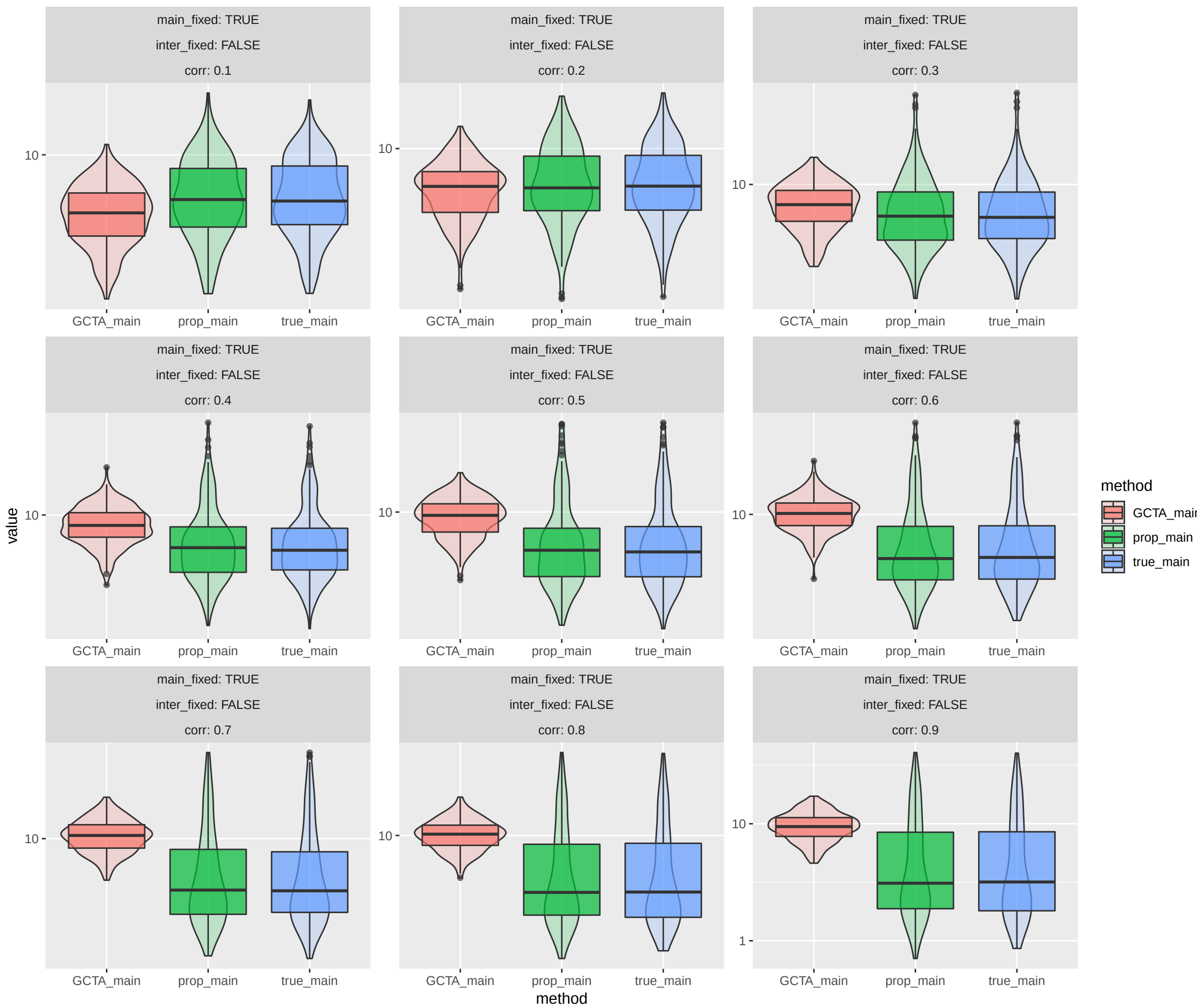

Figure 18: Total effect of chi fixed main and random interaction

5.4.3 β is random and Γ is random (Chi df = 1)

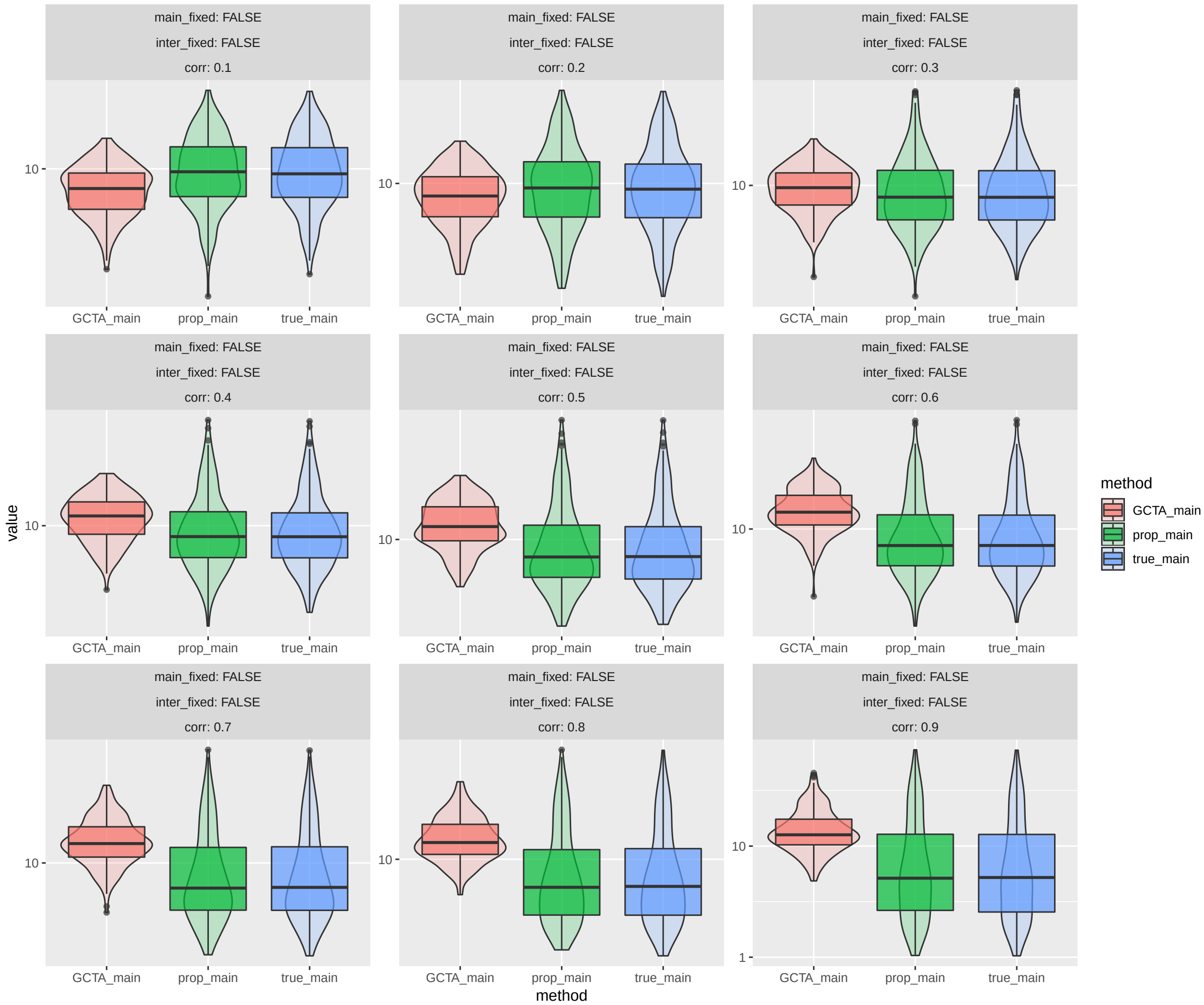

Figure 19: Total effect of chi random main and interaction

6 Conclusion

7 Further work