Supplementary Information for "A Regression Tree Method for Longitudinal and Clustered Data with Multivariate Responses"

Wenbo Jing[a] and Jeffrey S. Simonoff[a]

[a]New York University Leonard N. Stern School of Business, New York, NY, USA

ARTICLE HISTORY
Compiled February 24, 2023

A. Additional Results for Simpler Bivariate Tree

A.1. Object-Level PMSE for All of the Competitor Methods

We present the full results for all of the methods in Figures 1–4 here for completeness. It is clear from the results that the RE-EM methods always outperform the linear methods and the non-longitudinal tree method.

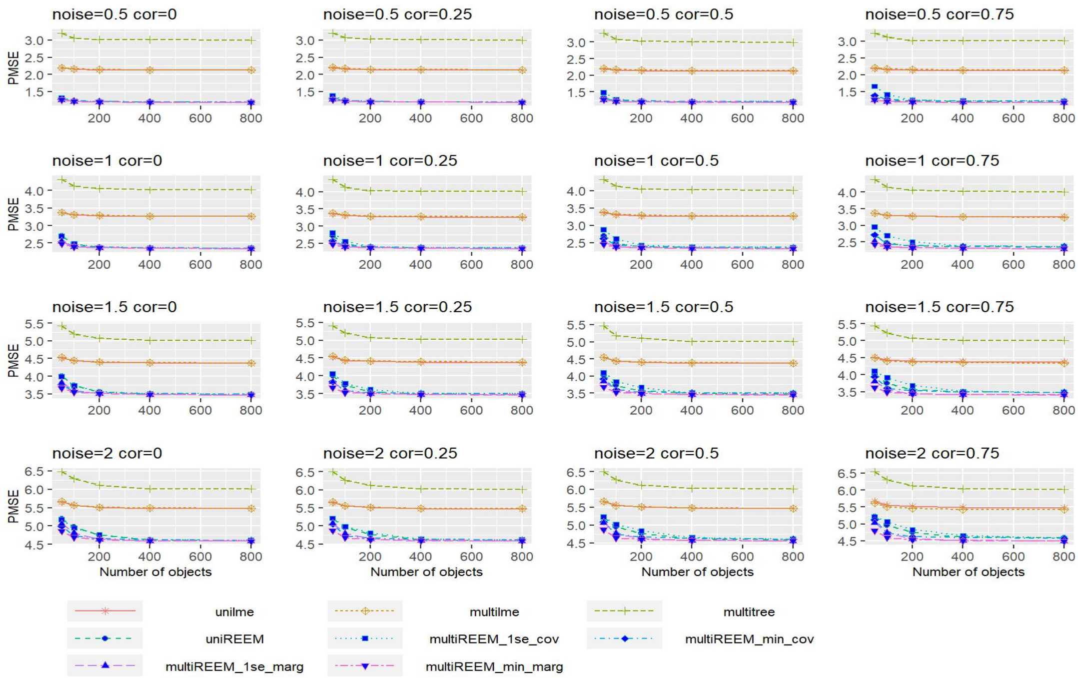

Figure 1. PMSE in object for $T_i = 5$ for the simpler bivariate tree structure

CONTACT Wenbo Jing. Email: wj2093@stern.nyu.edu

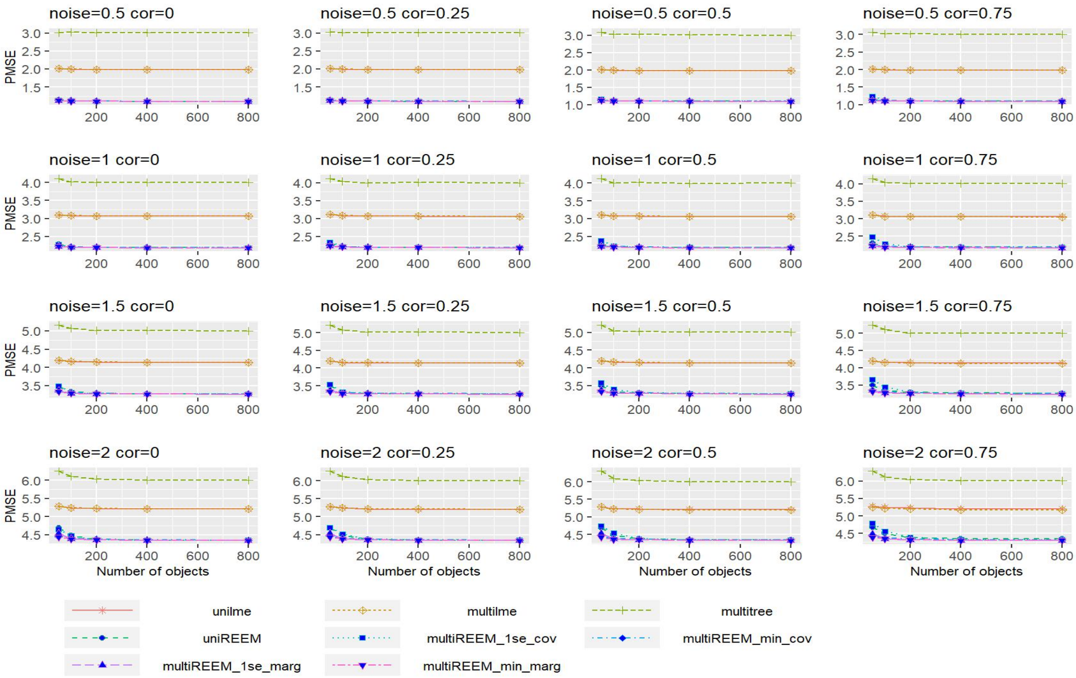

Figure 2. PMSE in object for $T_i = 10$ for the simpler bivariate tree structure

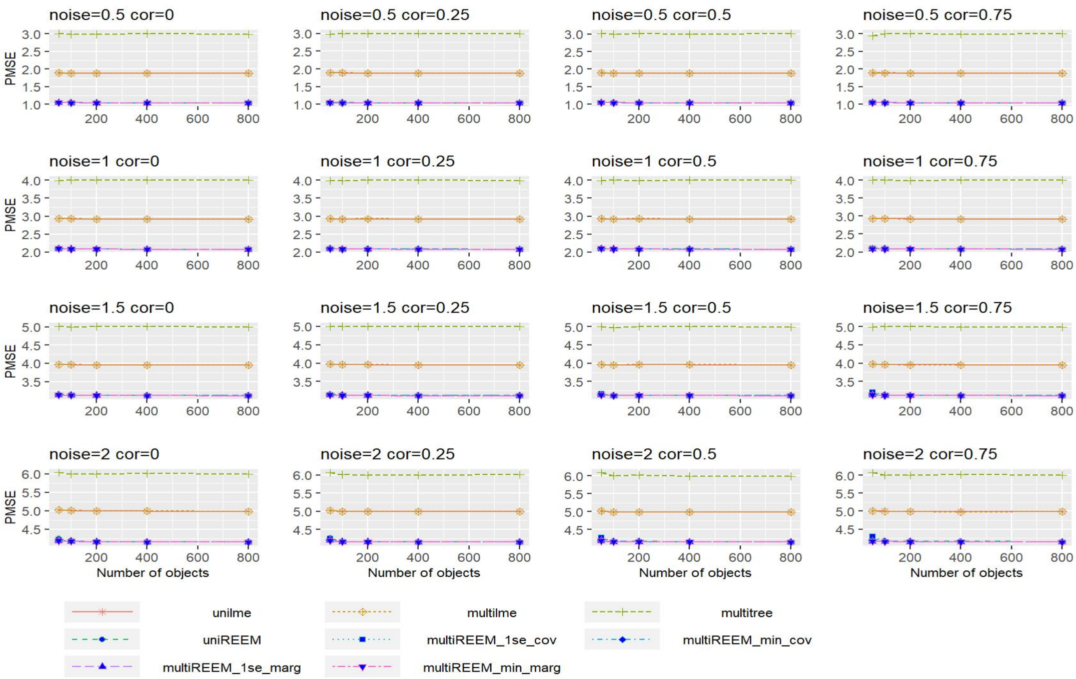

Figure 3. PMSE in object for $T_i = 25$ for the simpler bivariate tree structure

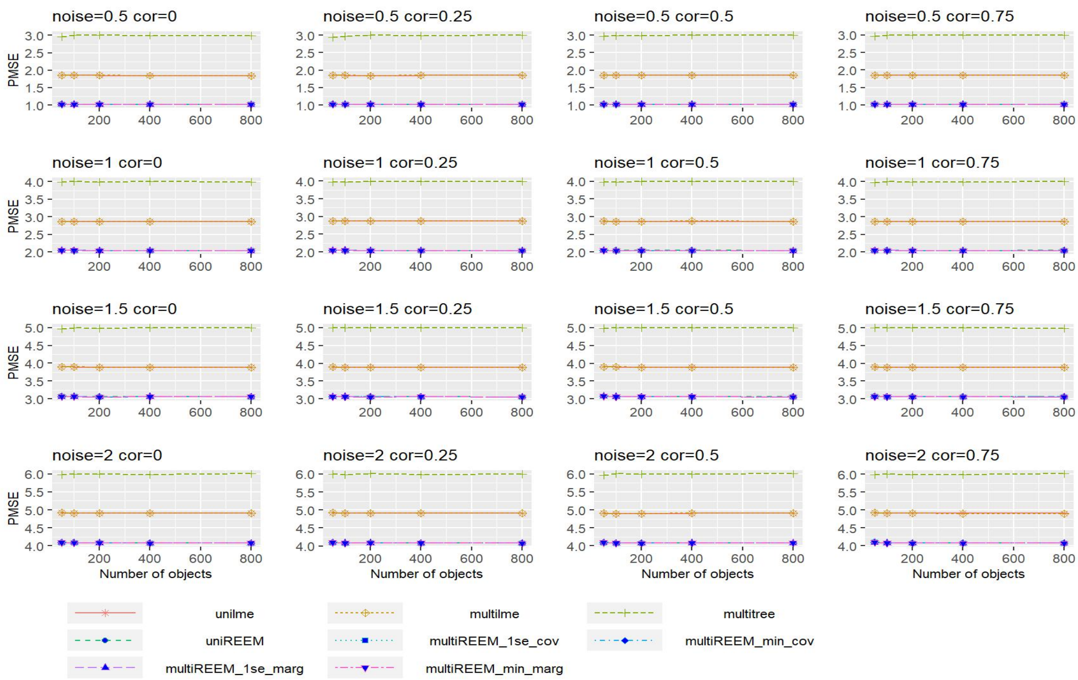

Figure 4. PMSE in object for $T_i = 50$ for the simpler bivariate tree structure

A.2. Fixed Effects Estimation and Random Effects Prediction

Figures 5–8 summarize the *estimation mean squared error* (EMSE) for the fixed effects and the *prediction mean squared error* (PMSE) for the random effects in the simpler bivariate tree setting. The Multivariate RE-EM tree method always estimates the fixed effects slightly better than "uniREEM," although their recovering rates are almost the same. This indicates that separate univariate RE-EM trees are able to recover the tree structure correctly but cannot estimate the true values on each terminal node as well as the Multivariate RE-EM tree. For the random effects, the Multivariate RE-EM tree also performs no worse than "uniREEM," and it significantly outperforms "uniREEM" as the noise or the correlation increases.

Additionally, we provide the EMSEs for estimating the correlation σ_{12} of the random effects in Figures 9–12. The competitors include the "multilme" method and all versions of "multiREEMtree" methods. The true value of σ_{12} is given by "cor" above each subfigure, and the average values of $(\widehat{\sigma}_{12} - \sigma_{12})^2$ over 100 independent runs are plotted. The estimation errors of Multivariate RE-EM tree with "marginal" standardization are always significantly lower than "multilme," and the Multivariate RE-EM tree with "covariance" standardization performs better when the true correlation is larger.

A.3. No-Random-Effect Simulations

In this section, we show the results of simulations with no random effects, i.e., $\boldsymbol{B} = 0$. Figures 13 and 14 present the EMSE of the fixed effect and the PMSE for object-level prediction, where the competitors are all of the tree-based methods. The MRT ("multitree") method and the multivariate RE-EM trees with "marginal" standardization

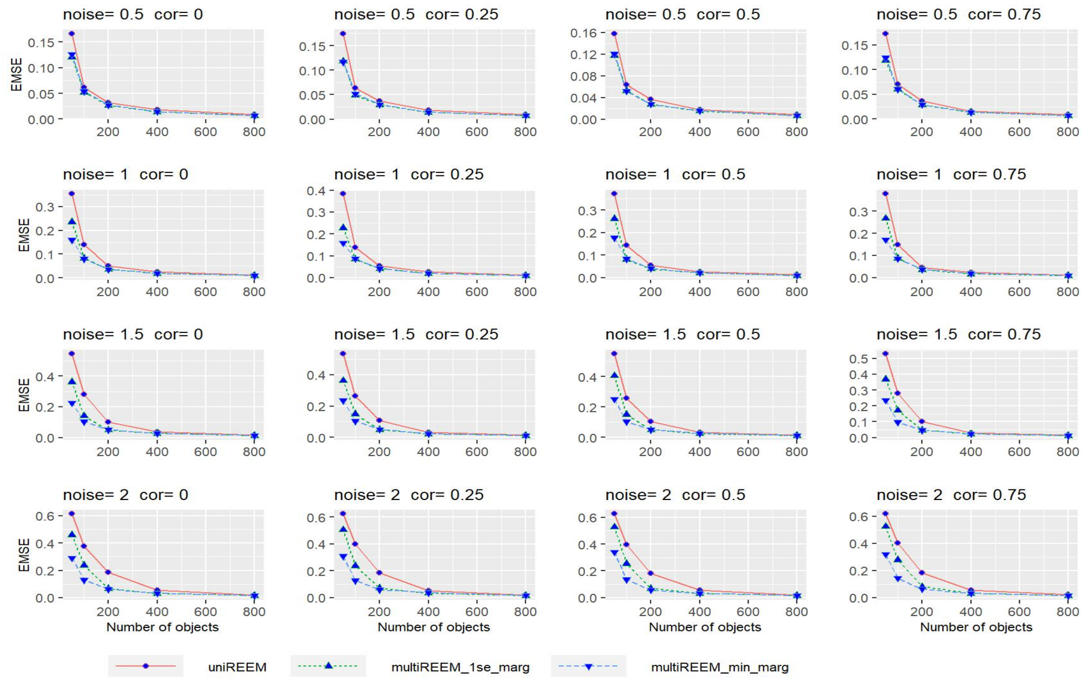

Figure 5. Estimation Mean Squared Error for fixed effects with $T_i = 5$

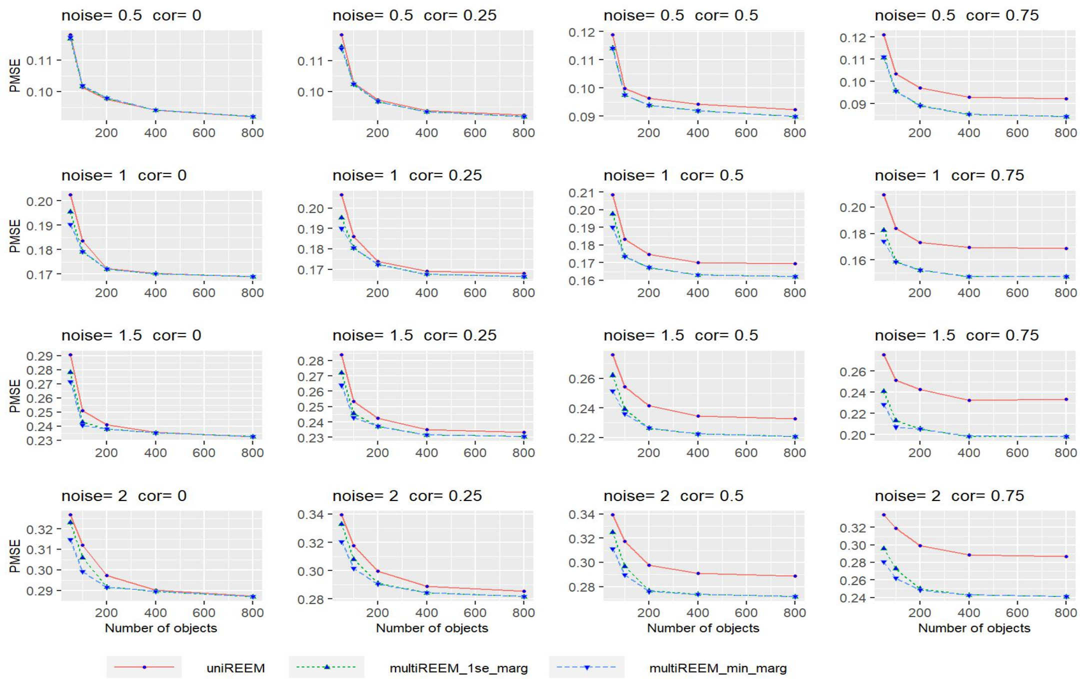

Figure 6. Prediction Mean Squared Error for random effects with $T_i = 5$

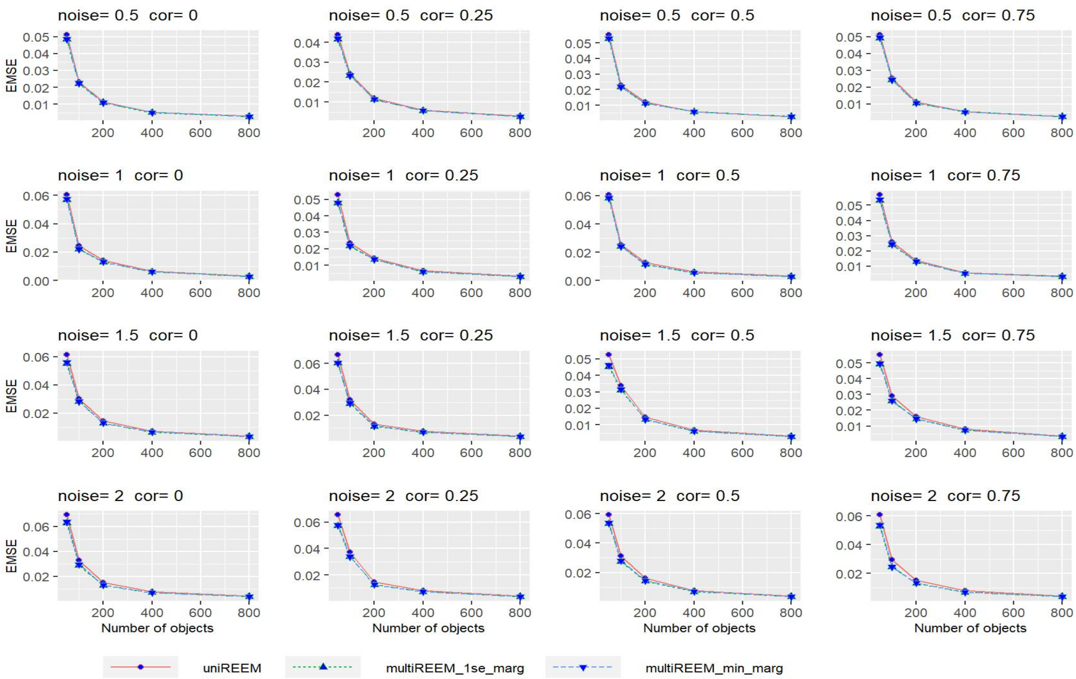

Figure 7. Estimation Mean Squared Error for fixed effects with $T_i = 50$

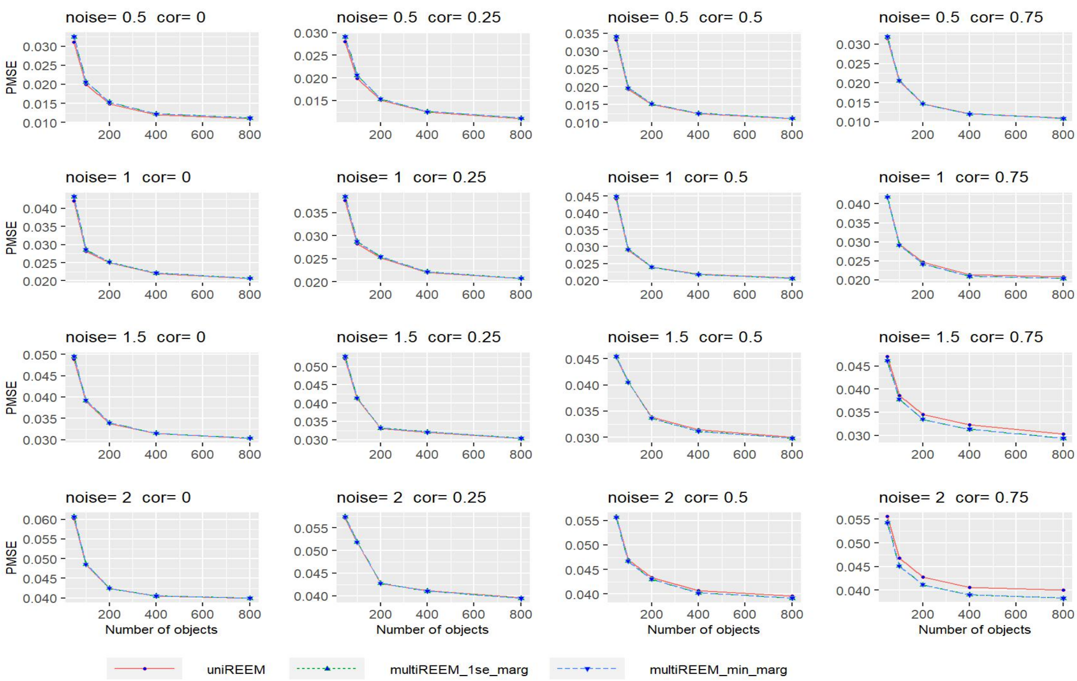

Figure 8. Prediction Mean Squared Error for random effects with $T_i = 50$

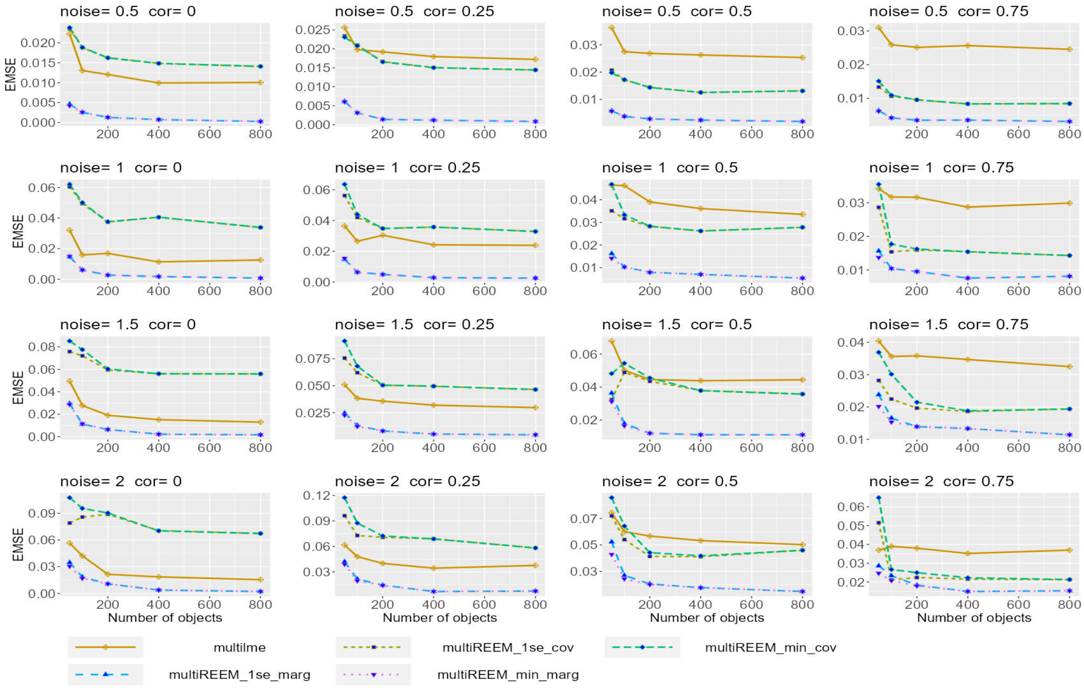

Figure 9. Estimation Mean Squared Error for the correlation σ_{12} of the random effects with $T_i = 5$

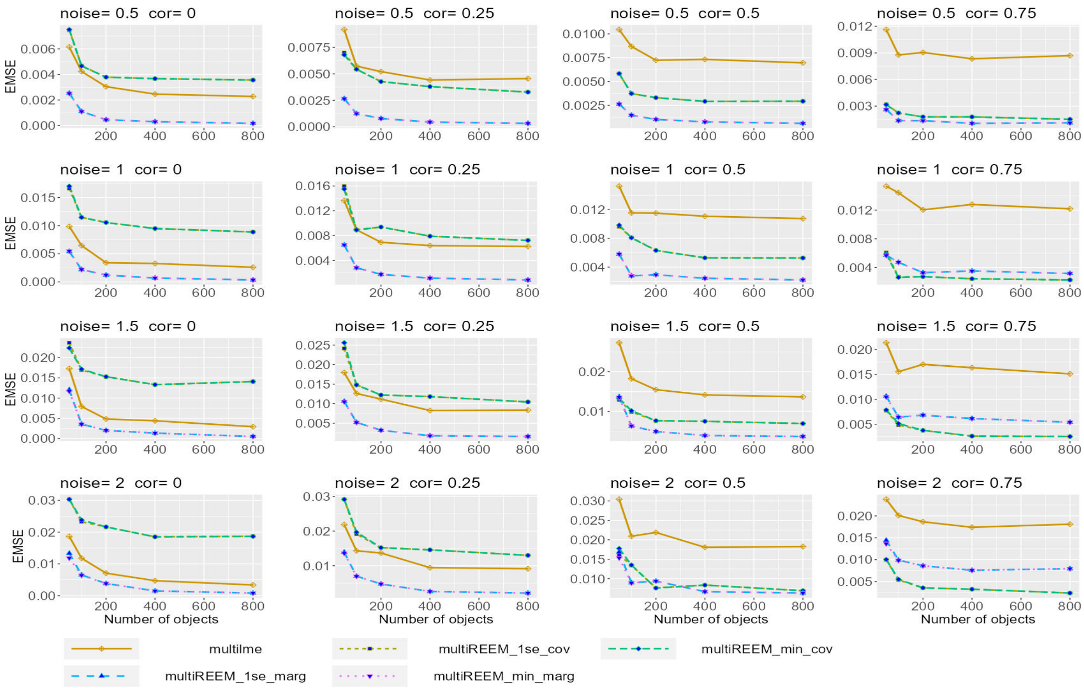

Figure 10. Estimation Mean Squared Error for the correlation σ_{12} of the random effects with $T_i = 10$

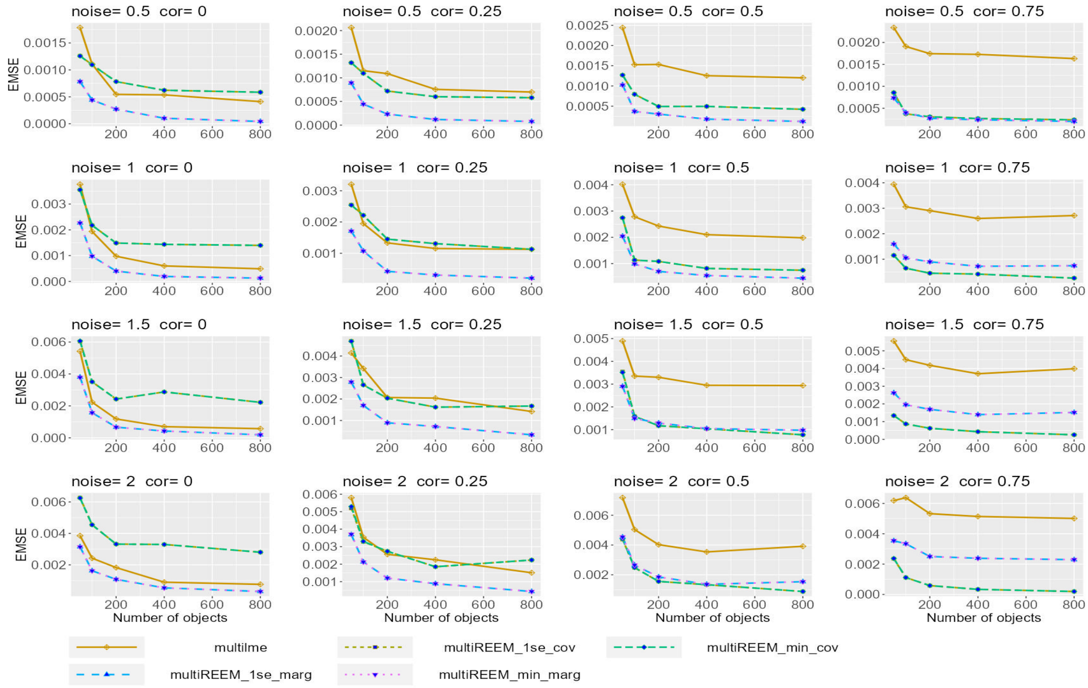

Figure 11. Estimation Mean Squared Error for the correlation σ_{12} of the random effects with $T_i = 25$

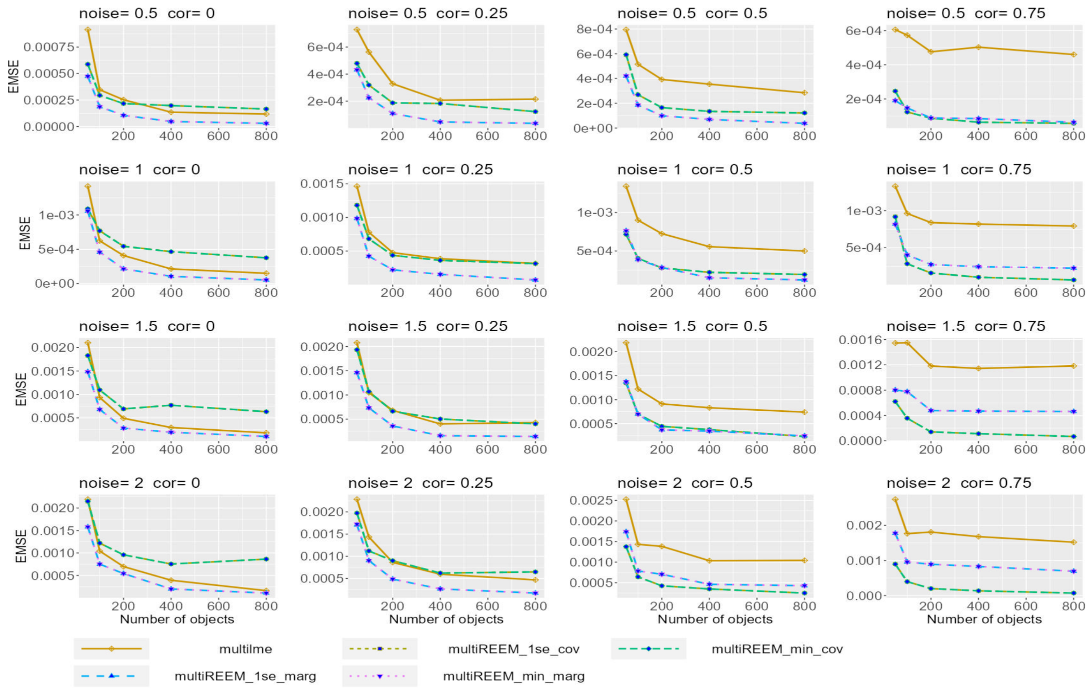

Figure 12. Estimation Mean Squared Error for the correlation σ_{12} of the random effects with $T_i = 50$

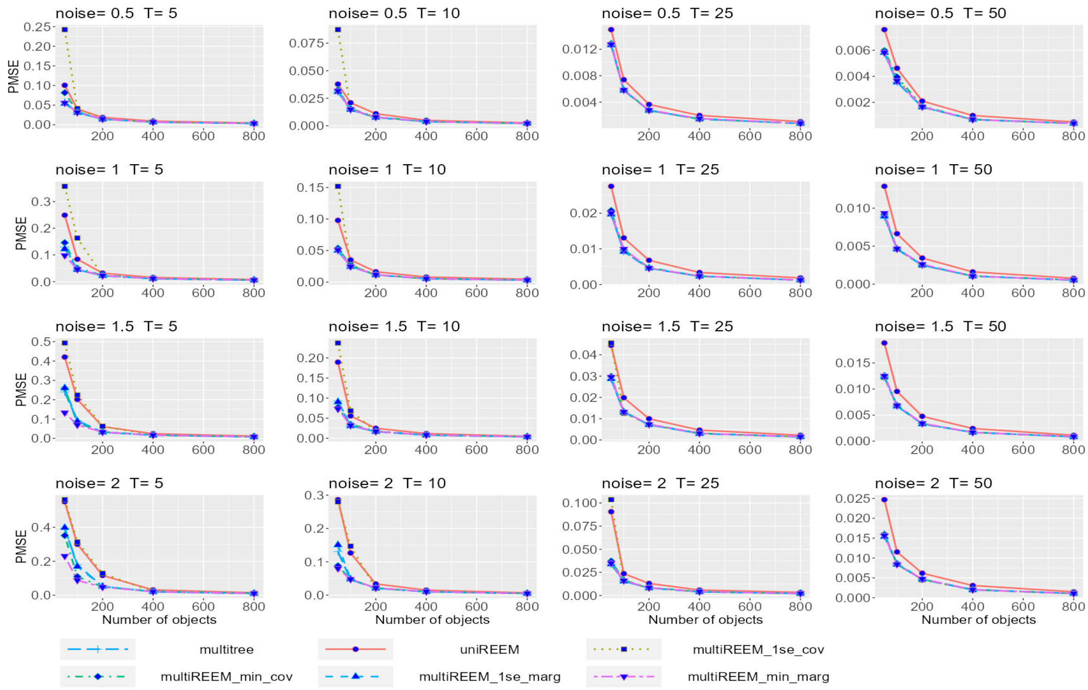

Figure 13. EMSE of the fixed effect for the simulations with no random effect

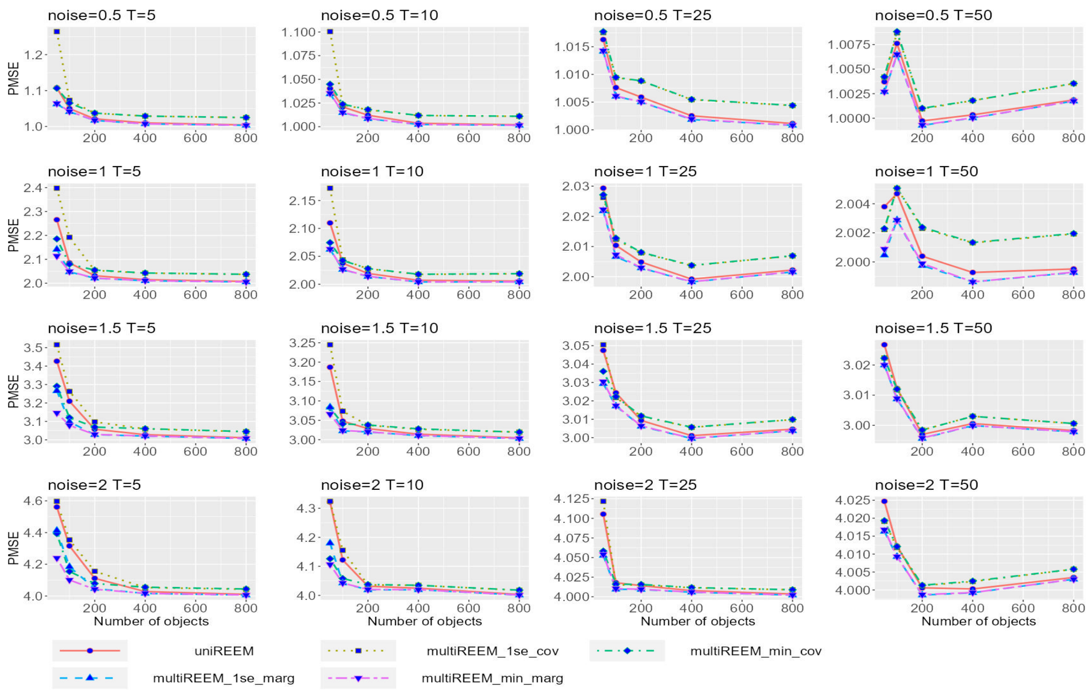

Figure 14. PMSE in object for the simulations with no random effect

always outperform other method, with no meaningful difference between them. This verifies that the proposed "multiREEMtree" algorithm works as well as an MRT in no-random-effect cases.

B. Additional Results for the Complex Bivariate Tree Structure

The PMSEs and the tree recovering rates for $T_i = 10, 25, 50$ with the complex bivariate tree structure are shown in Figures 15–20.

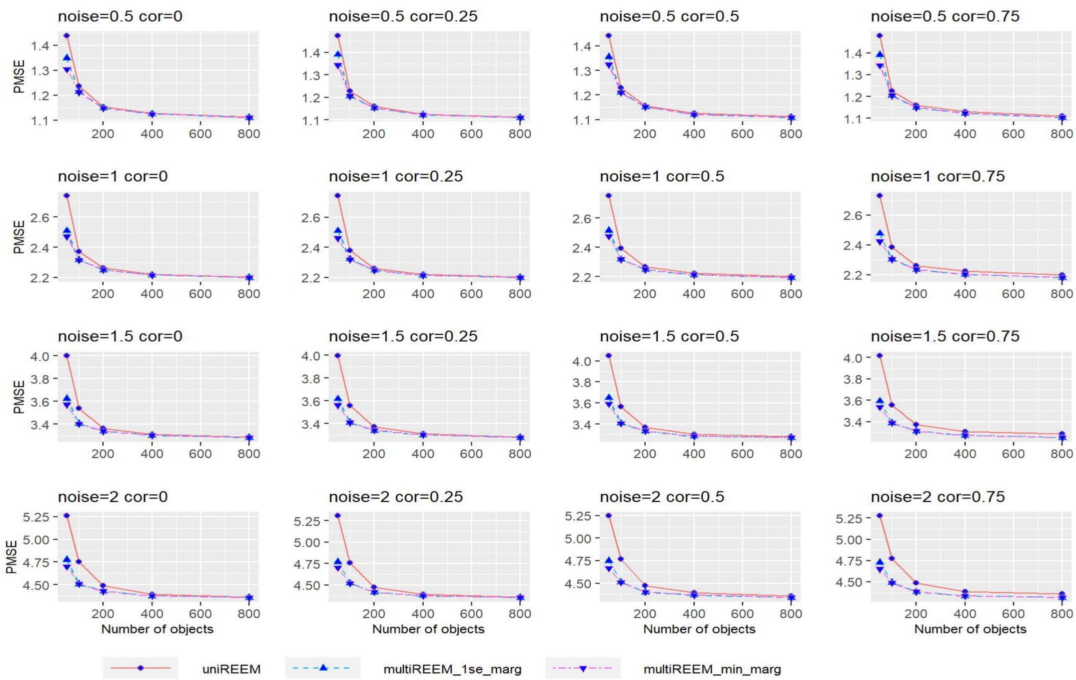

Figure 15. PMSE in object for $T_i = 10$ for the complex bivariate tree

C. Additional Results for Multivariate Tree with Five Response Variables

The PMSEs and the tree recovering rates for $T_i = 10, 25, 50$ with the five-response tree structure are shown in Figures 21–26.

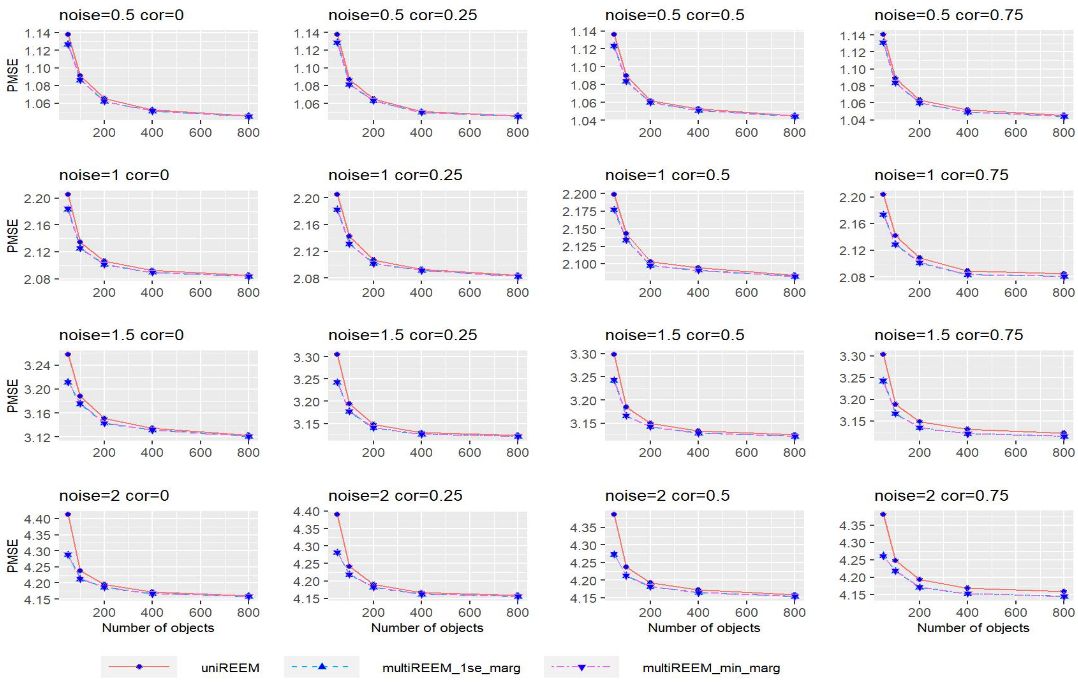

Figure 16. PMSE in object for $T_i = 25$ for the complex bivariate tree

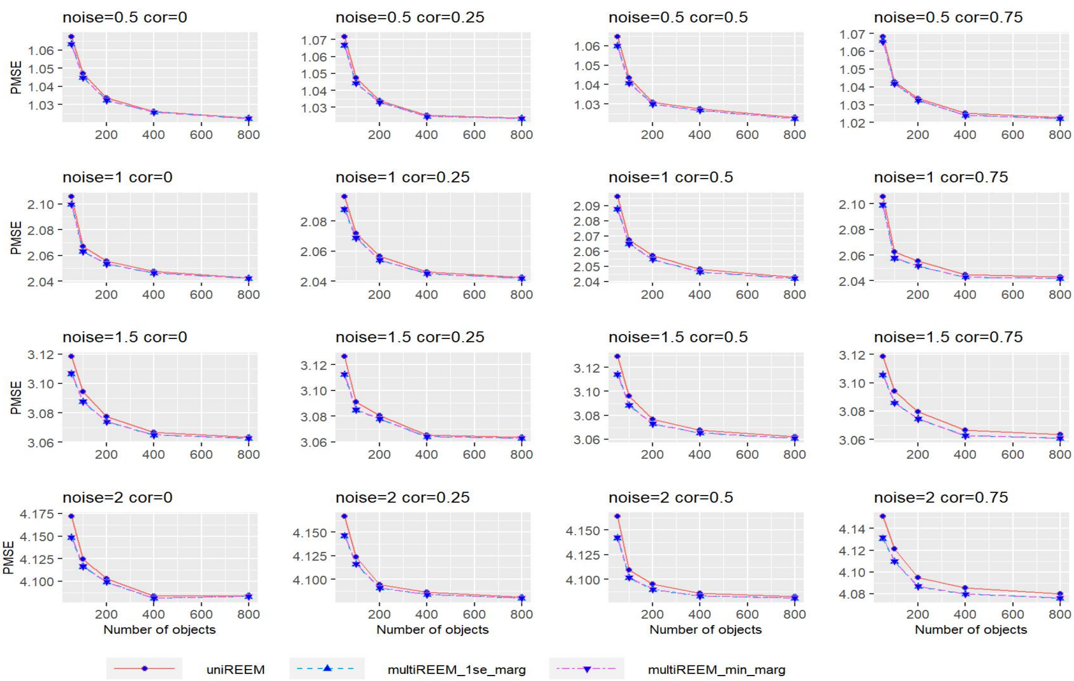

Figure 17. PMSE in object for $T_i = 50$ for the complex bivariate tree

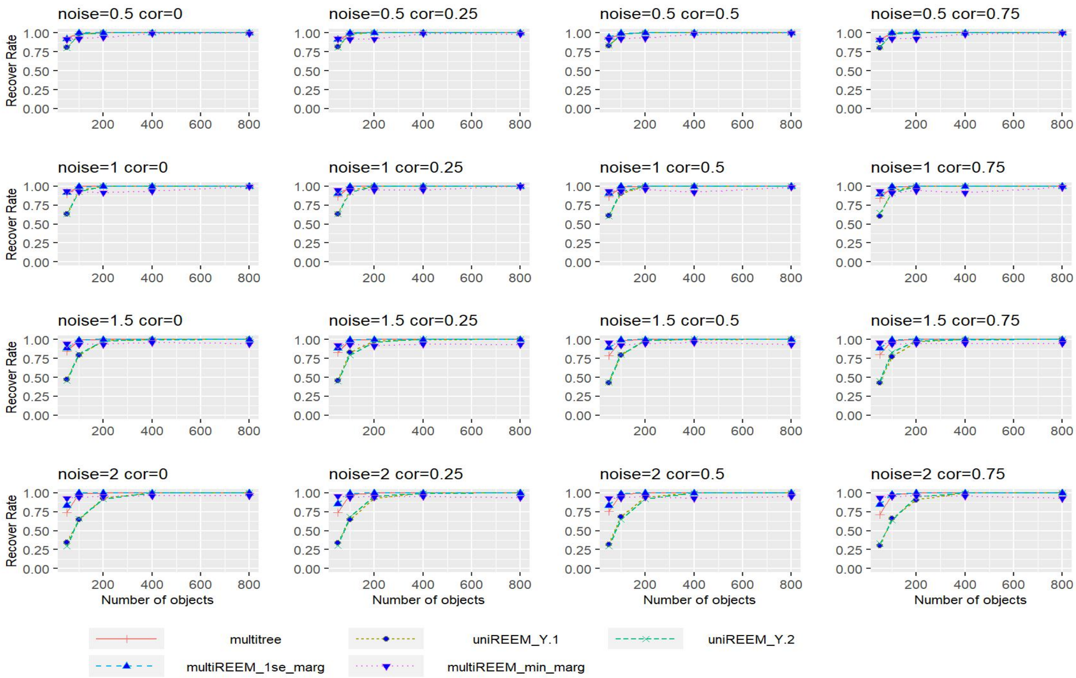

Figure 18. Recovering rate for $T_i = 10$ for the complex bivariate tree

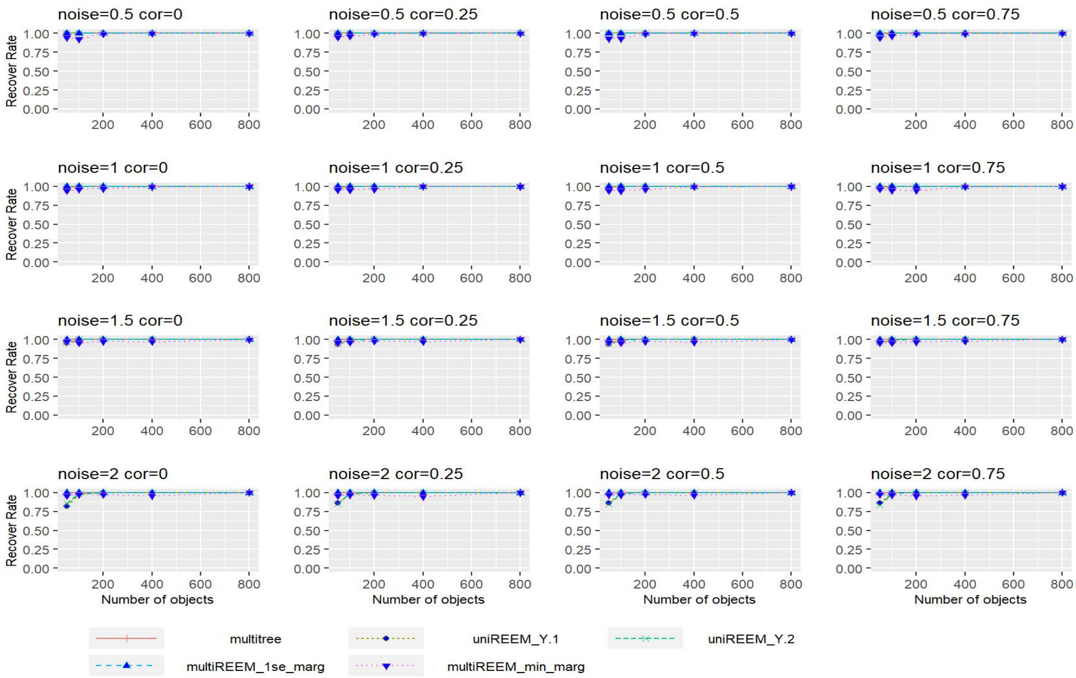

Figure 19. Recovering rate for $T_i = 25$ for the complex bivariate tree

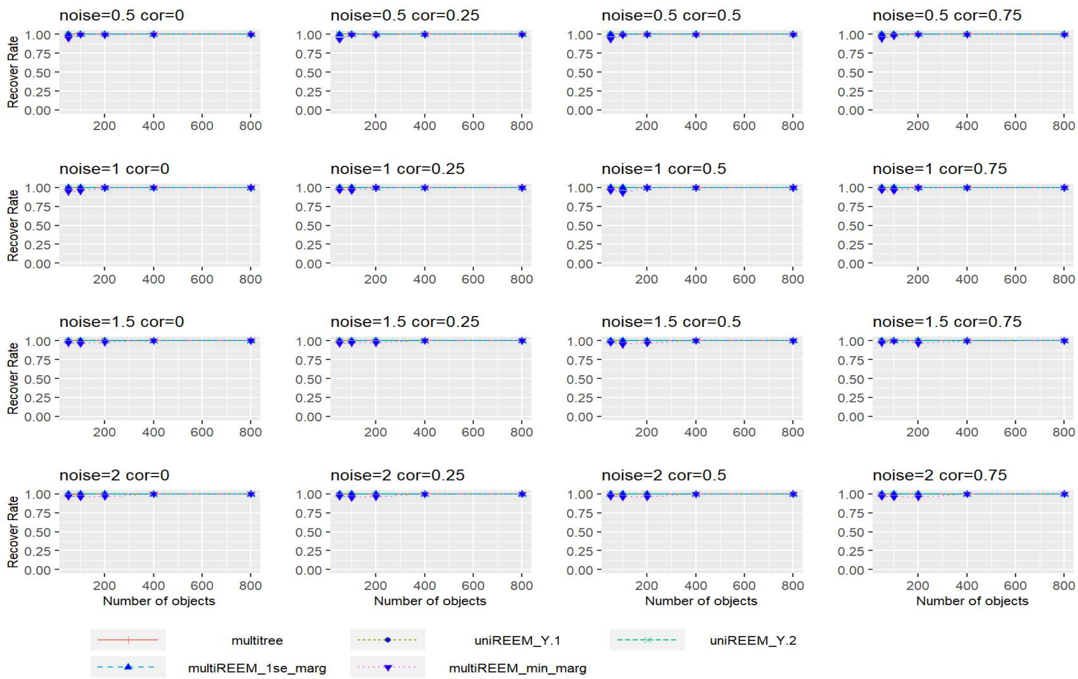

Figure 20. Recovering rate for $T_i = 50$ for the complex bivariate tree

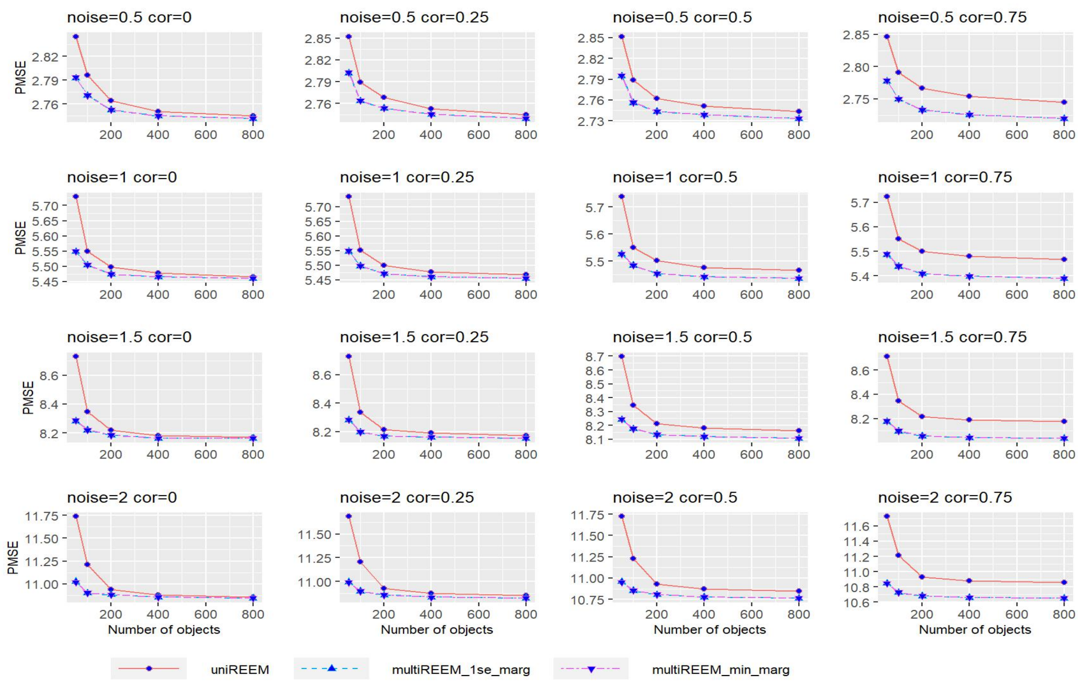

Figure 21. PMSE in object for $T_i = 10$ for multivariate tree with five response variables

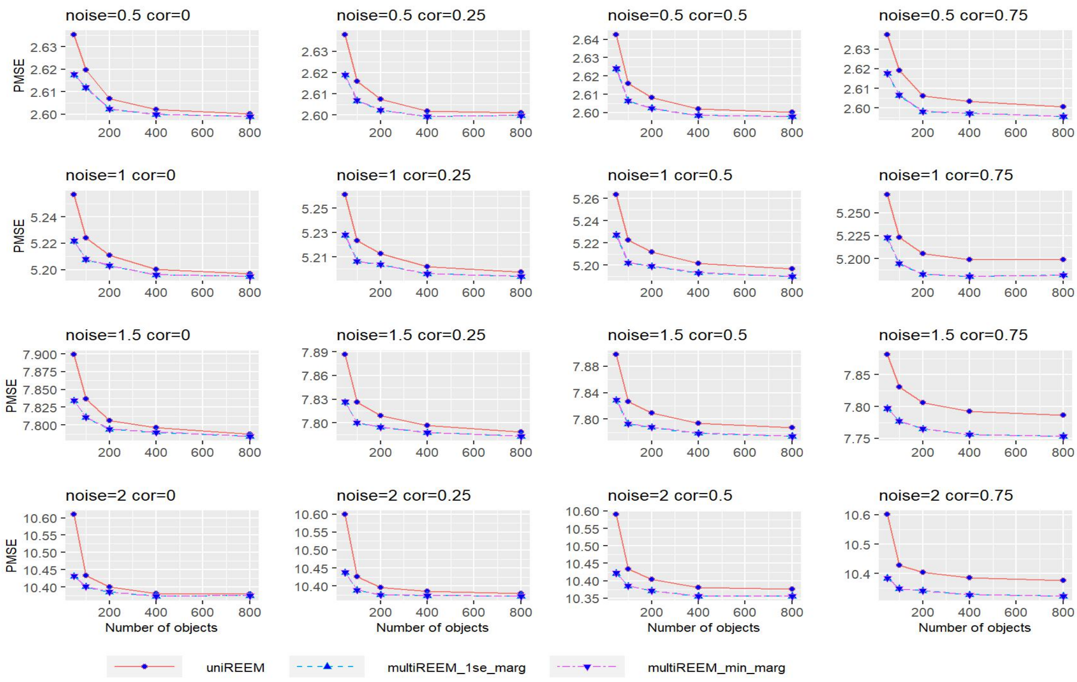

Figure 22. PMSE in object for $T_i = 25$ multivariate tree with five response variables

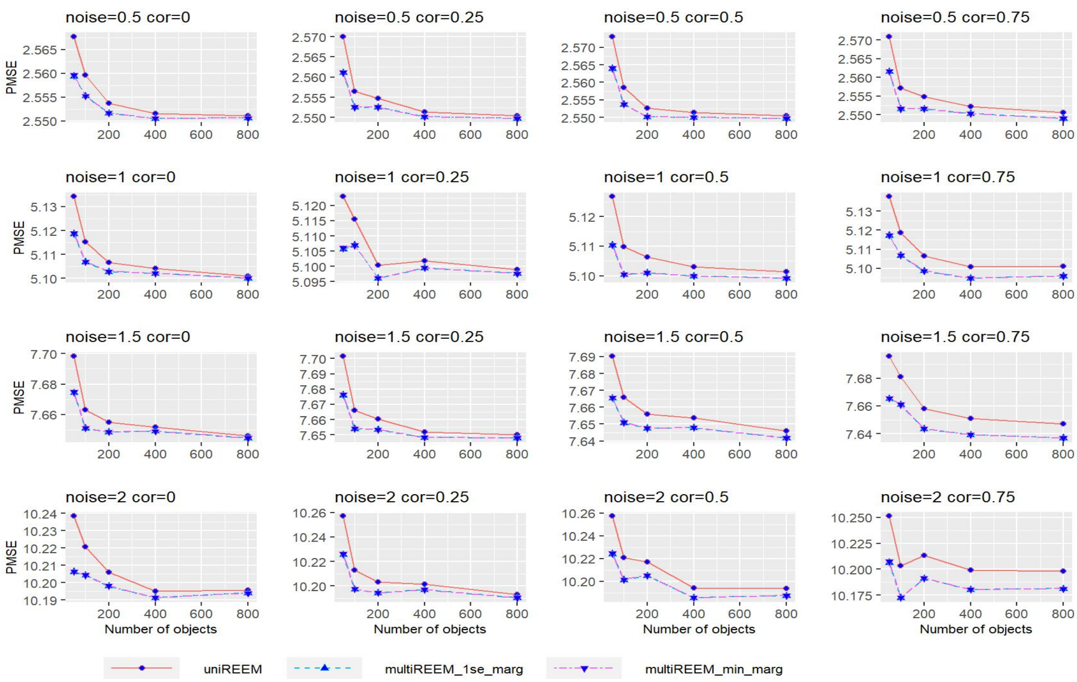

Figure 23. PMSE in object for $T_i = 50$ multivariate tree with five response variables

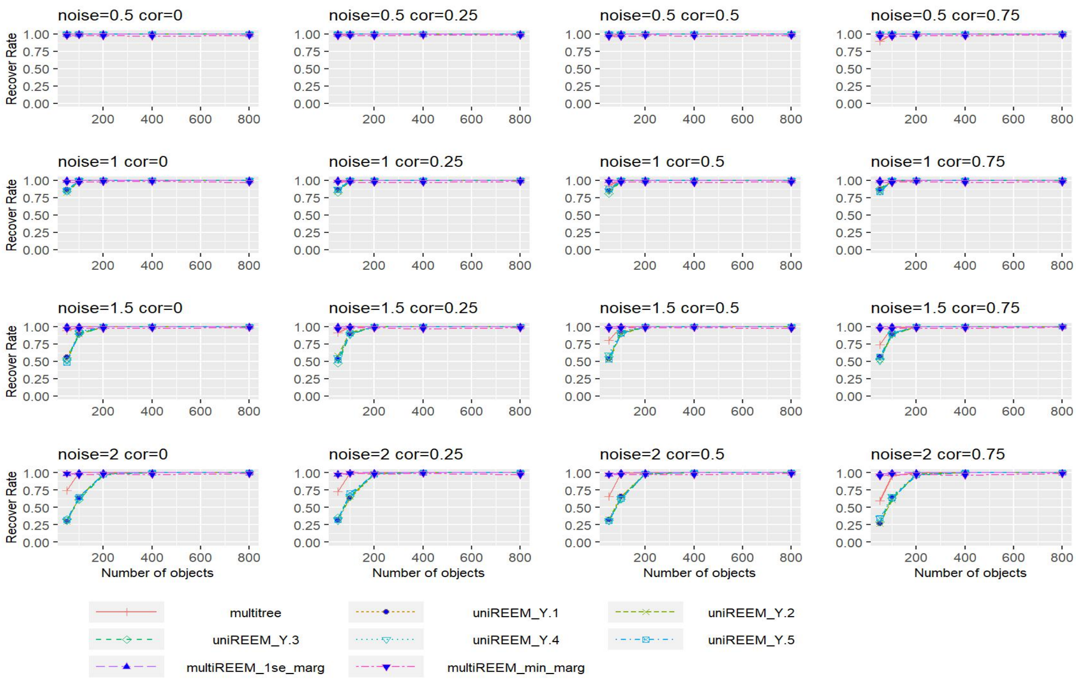

Figure 24. Recovering rate for $T_i = 10$ multivariate tree with five response variables

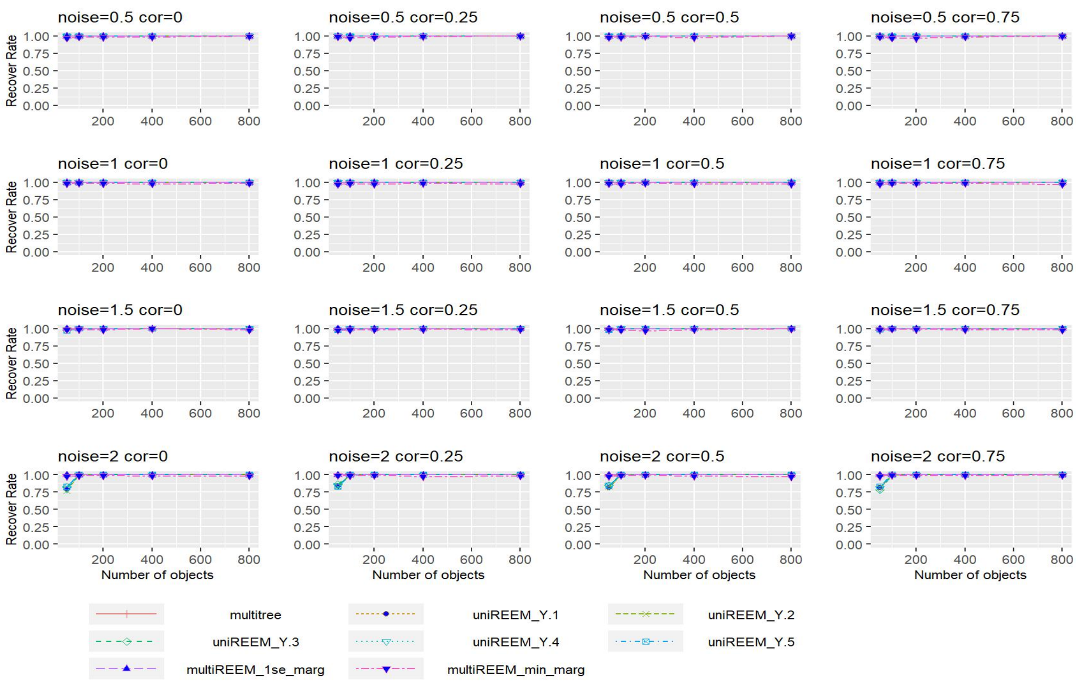

Figure 25. Recovering rate for $T_i = 25$ multivariate tree with five response variables

Figure 26. Recovering rate for $T_i = 50$ multivariate tree with five response variables