Supplementary Material for: Survival Forests for Data with Dependent Censoring

Hoora Moradian, Denis Larocque* and François Bellavance
Department of Decision Sciences, HEC Montréal, Canada

*Corresponding author. Department of Decision Sciences, HEC Montréal, 3000 chemin de la Côte–Sainte–Catherine, Montréal (Québec), Canada, H3T 2A7. E-mail:denis.larocque@hec.ca

The code to perform the simulations and the run the data example is available at:
http://neumann.hec.ca/pages/denis.larocque/Personnel/documents/code.zip

1 Detailed simulation results

The discussion in the paper is sufficient to get the general ideas about the results and to highlight the meaningful findings. But for completeness, Figures 1 to 27 present the detailed results for all 45 scenarios investigated in the article. The results for each combination (9 in all) of DGP (1, 2 or 3) and proportion of censoring (20%, 40% or 60%) is summarized in three consecutive figures. For each combination, the first figure present the plots when true τ is 0 and 0.2. The second one is when true τ is 0.4 and 0.6. The third one is different. The upper plot present the case of a true τ of 0.8. The lower plot in this third figure is different. The plot is divided in 5 sections according to the value of true τ. Within a section (a given value of true τ), only the best one (i.e. the best working τ) is presented for each family (CF, RSF, L1, L1S, pf5, pf9) of methods. This allows an easy comparison of the best results, i.e. what each family can achieve if we know the best working τ, across all families.

To fix ideas, we begin with Figures 1 to 3 that correspond to the combination DGP 1 and 20% of censoring. When true τ is 0 (upper plot in Figure 1), then the CF method is slightly better (lowest median IAE over the 100 runs) than the others, pf5 being a close second. The best value of working τ is 0.2 for all methods. The original RSF does not perform well in this scenario. When true τ is 0.2 (lower plot in Figure 1), the pf5 method is slightly better (CF being a close second). This time the best value of working τ is either 0.2 or 0.4 depending on the method. When true τ is 0.4 (upper plot in Figure 2), pf5 is again slightly better followed closely by CF. We also begin to see a drift in patterns as higher values of working τ are preferable. Similar patterns occur when true τ is 0.6 (lower plot in Figure 2) and the drift towards having better performances with higher working τ values continues. When true τ is 0.8 (upper plot in Figure 3), pf5 continues to have a slight edge and now the best choice of working τ is 0.8 except maybe for CF where 0.6 and 0.8 offer similar performances. We see also that using a too small value of working τ, especially 0 and 0.2, degrades the performance

severely. The lower plot in Figure 3 compares the best choice of working τ within each family for each value of true τ. As mentioned above, we see that CF and pf5 are the methods with the best potential (if we could pick the right working τ). Indeed, CF can achieve the best result when true τ is 0 and pf5 can achieve the best result for the other values of true τ. The fact that the L1 and L1S methods offer similar performances can be seen scenario by scenario here. Consequently, modifying the splitting rule does not seem necessary once an adequate final estimator is used.

Without going into details, similar patterns occur in Figures 4 to 6 (DGP 1 and 20% of censoring) and 7 to 9 (DGP 1 and 60% of censoring). But we can also see that a good choice of working τ becomes more important as the proportion of censoring increases. Indeed, for a given true τ, the difference between the best and worst performances within a family increases as the proportion of censoring increase. This is intuitively expected since there is less information available when the proportion of censoring is high and thus obtaining a good performance relies more on an adequate choice of working τ. This is even clearer with DGP 2 (Figures 10 to 18). With 20% of censoring, the performances are similar within each family and even across families, with the exception of RSF-OR which has a worst performance. But when the proportion of censoring increases to 40% and 60%, the variability within a family increases and also the variability between the families.

With DGP 3 (Figures 19 to 27), the differences between the methods is greater right from the start with 20% censoring. This time, the CF method is considerably worse than the other methods, especially with 20% and 40% of censoring. Also different from the first two DGPs is the fact that the original RSF method, RSF-OR, is almost always better than RSF-KM, and sometimes by a large margin.

The last two figures (28 and 29) are similar to Figure 4 in the article but for DGP 2 and 3. They compare the results when the true copula is the Frank copula but the methods use Clayton as the working copula. Again we see that the results are not very sensitive to the choice of working copula.

To summarize, the same general conclusions as in the article can be made by looking at

the detailed results. All methods can perform well and the relative performance depends on the DGP. Modifying the splitting rule does not improve the results once an appropriate final estimator is used. The main driver of the performance is the choice of working τ. This choice becomes more important as the true τ increases and/or as the proportion of censoring increases. The RSF-KM method generally performed better than the original RSF-OR method.

2 Additional figures for the data example

As explained in the article, the PBC data set was divided in a training sample of size 300 and a sample of 12 "new" observations. The detailed results are provided for three typical subjects in the article to illustrate how the sensitivity to the choice of the dependency parameter varies from subject to subject. They were selected to illustrate cases of low, medium and high sensitivity. Figures 30 to 38 give the estimated survival functions for the nine other subjects. Different patterns and degrees of sensitivity can be seen. However, as noted in the article, the tendency is that the subjects with the worse prognoses (lower median survival times) have lower sensitivities and vice-versa. For completeness, these nine plots correspond to the following subjects (lines) in the PBC data set: 291, 225, 253, 48, 86, 249, 280, 47, 179. The three subjects in the article correspond to the lines 293, 4, and 169.

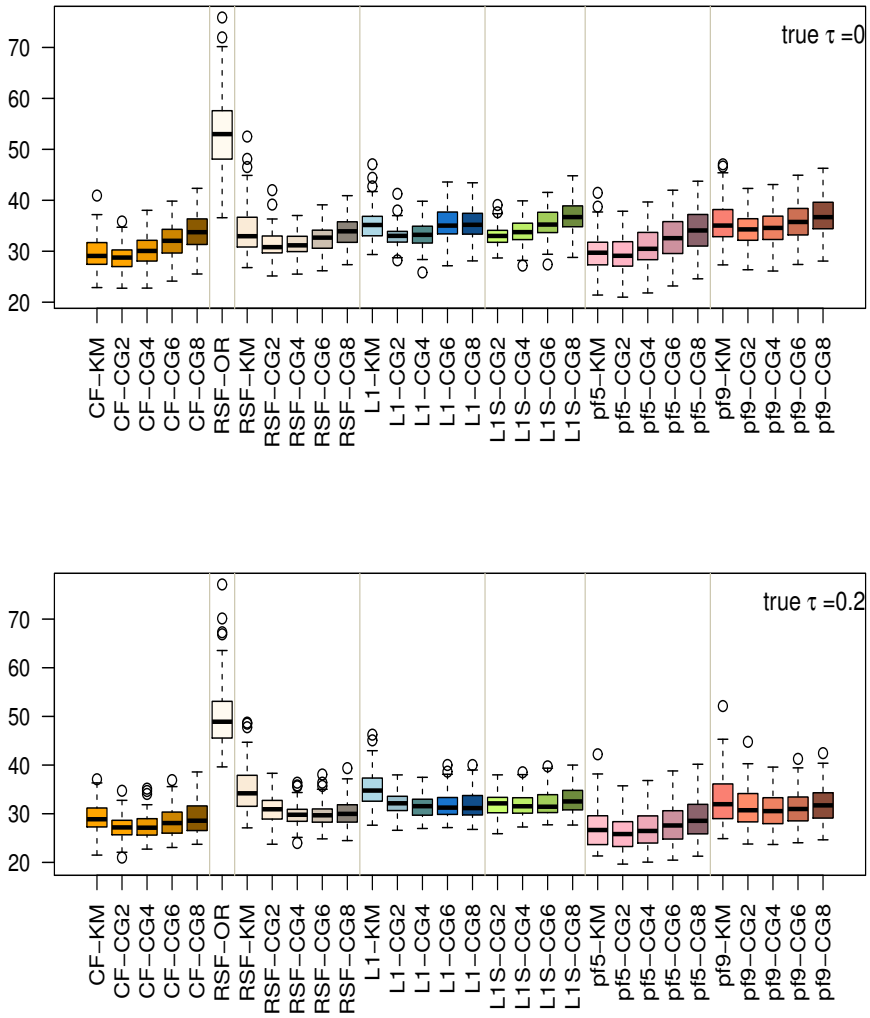

Figure 1: IAE of methods for DGP 1 at 20% censoring proportion with true τ values of 0 and 0.2

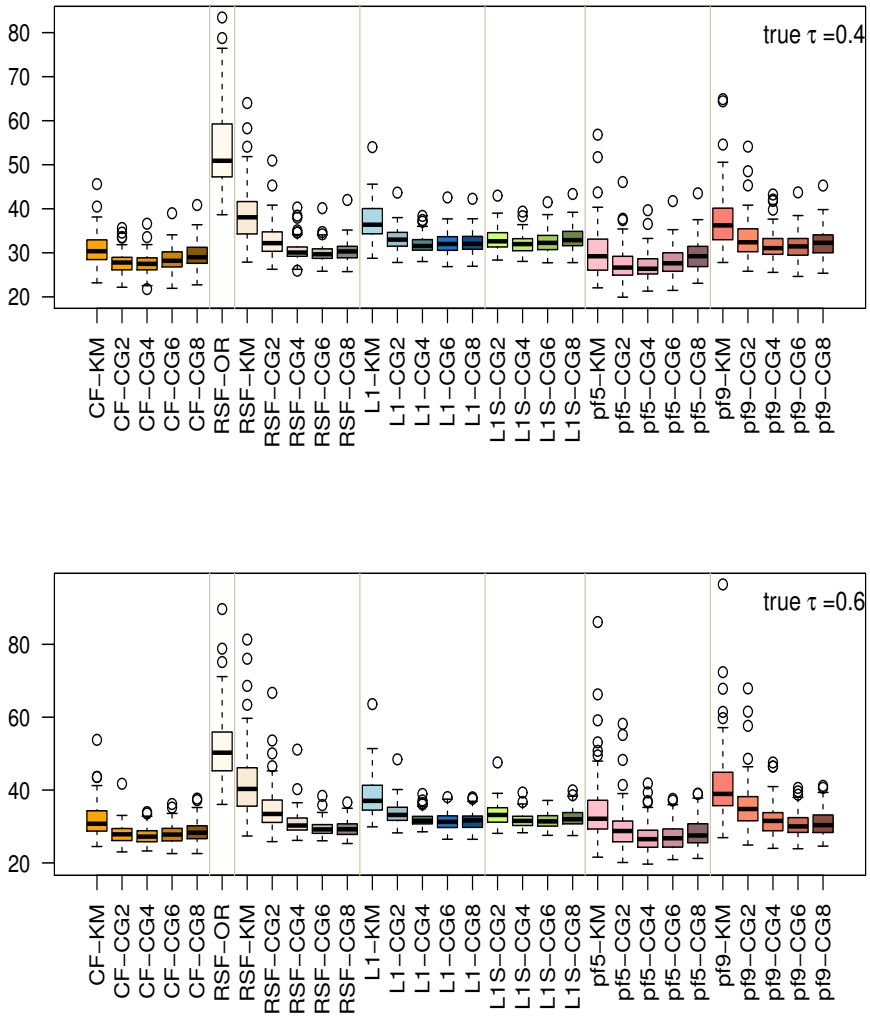

Figure 2: IAE of methods for DGP 1 at 20% censoring proportion with true τ values of 0.4 and 0.6

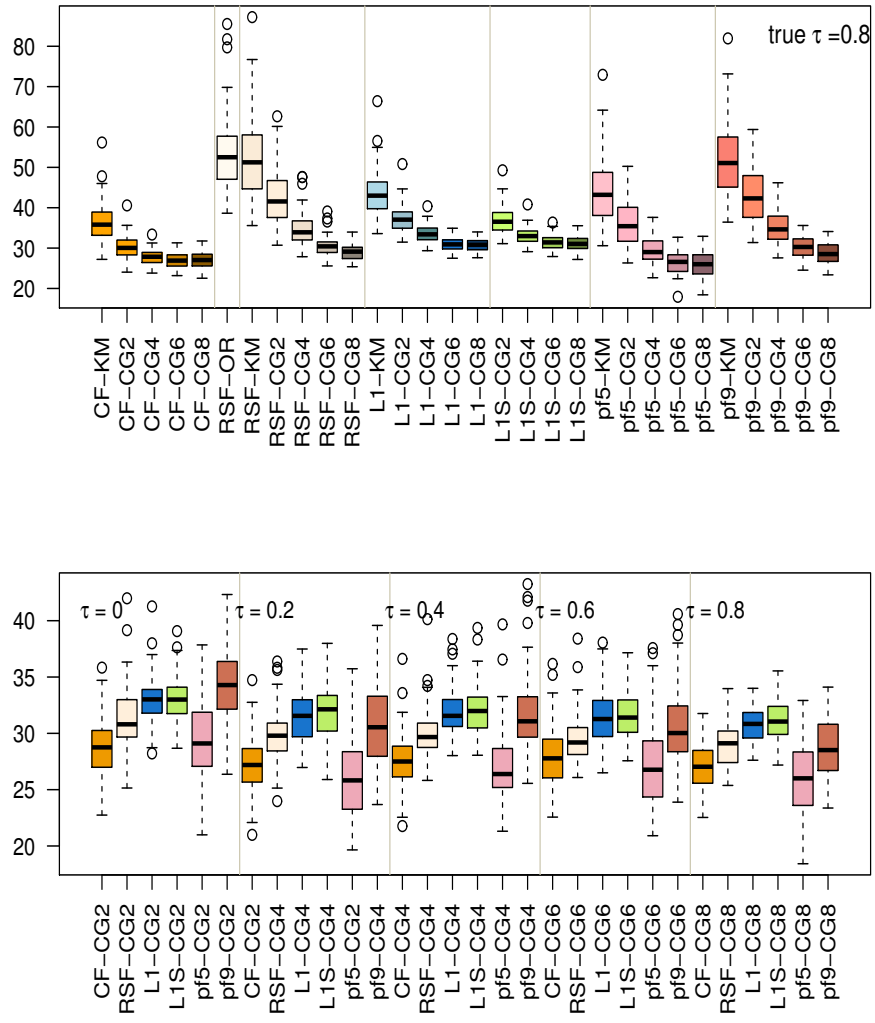

Figure 3: IAE of methods for DGP 1 at 20% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

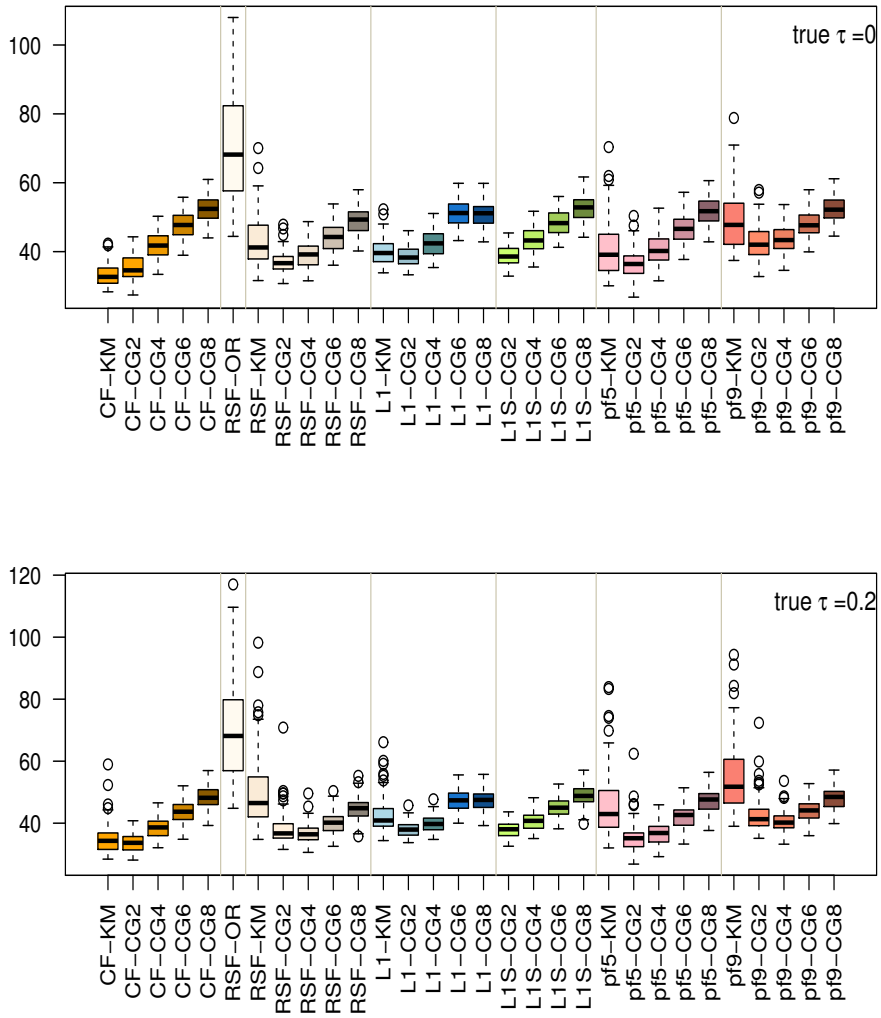

Figure 4: IAE of methods for DGP 1 at 40% censoring proportion with true τ values of 0 and 0.2

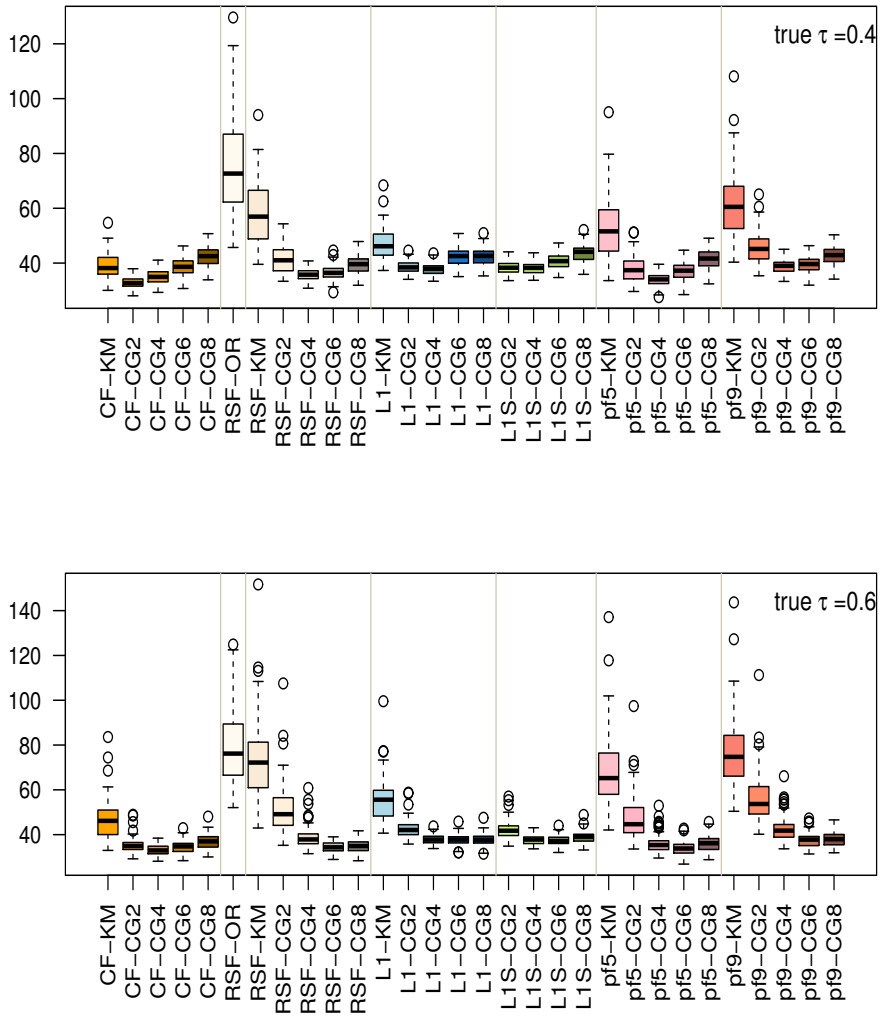

Figure 5: IAE of methods for DGP 1 at 40% censoring proportion with true τ values of 0.4 and 0.6

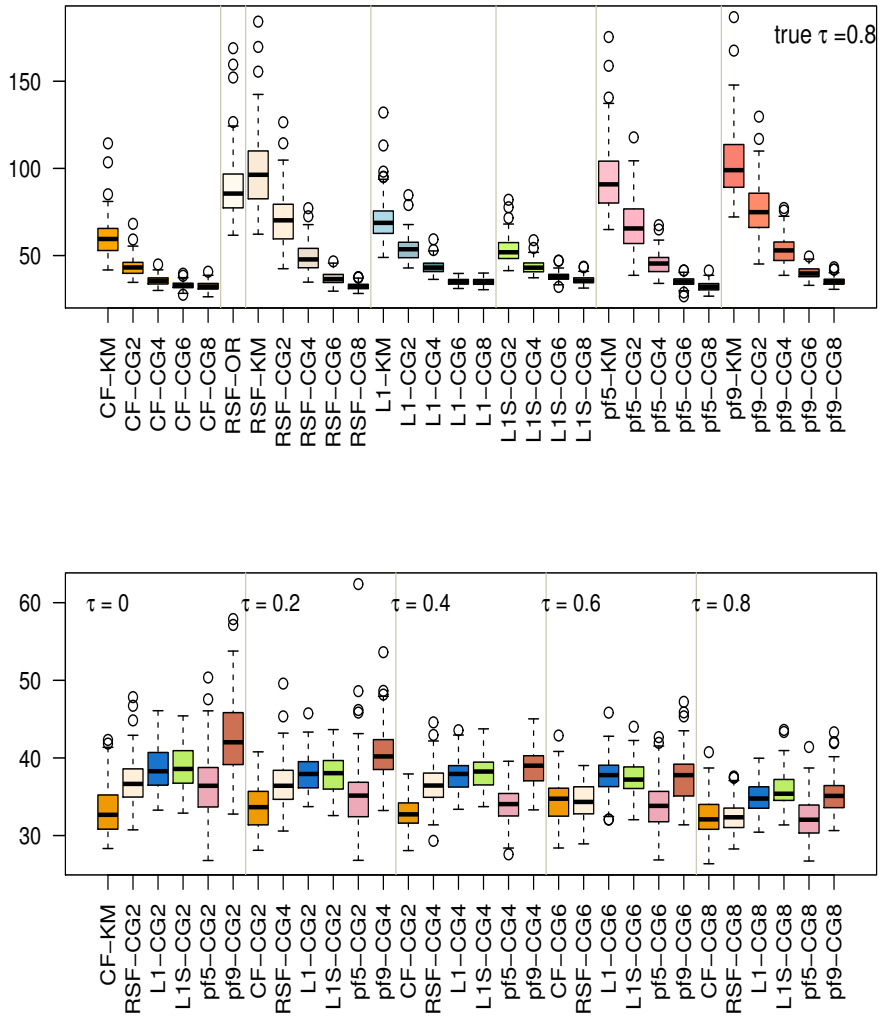

Figure 6: IAE of methods for DGP 1 at 40% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

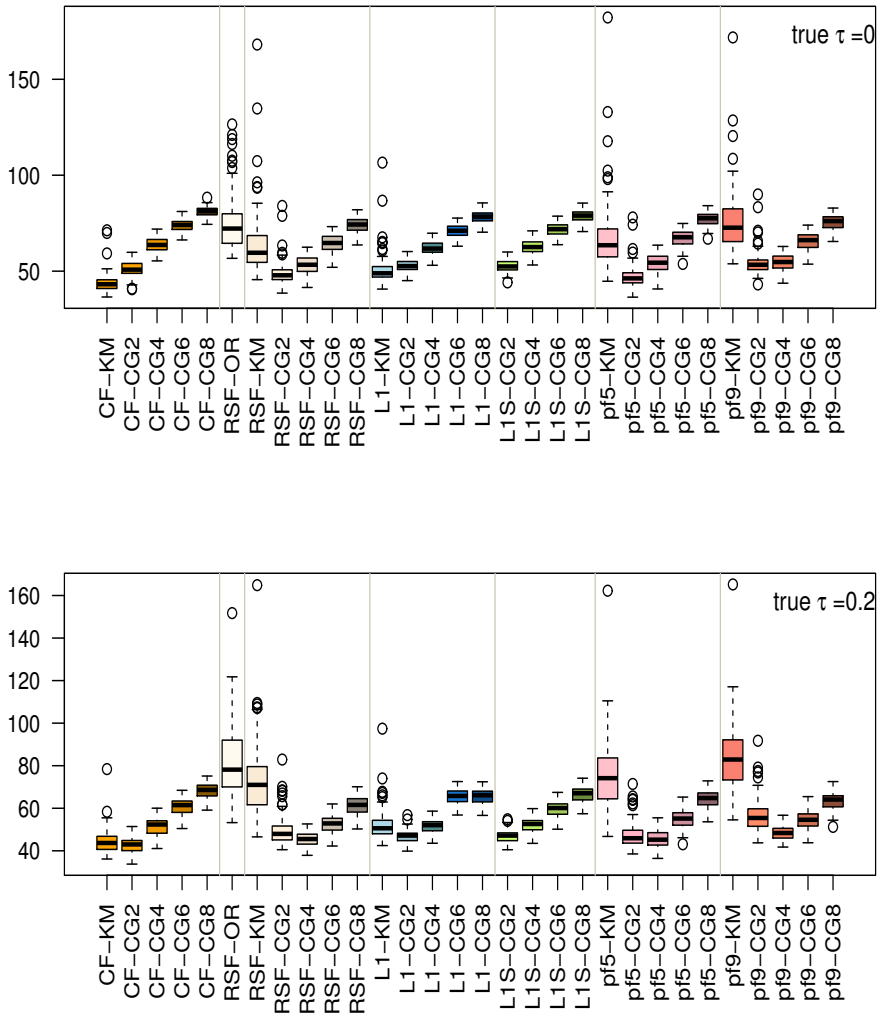

Figure 7: IAE of methods for DGP 1 at 60% censoring proportion with true τ values of 0 and 0.2

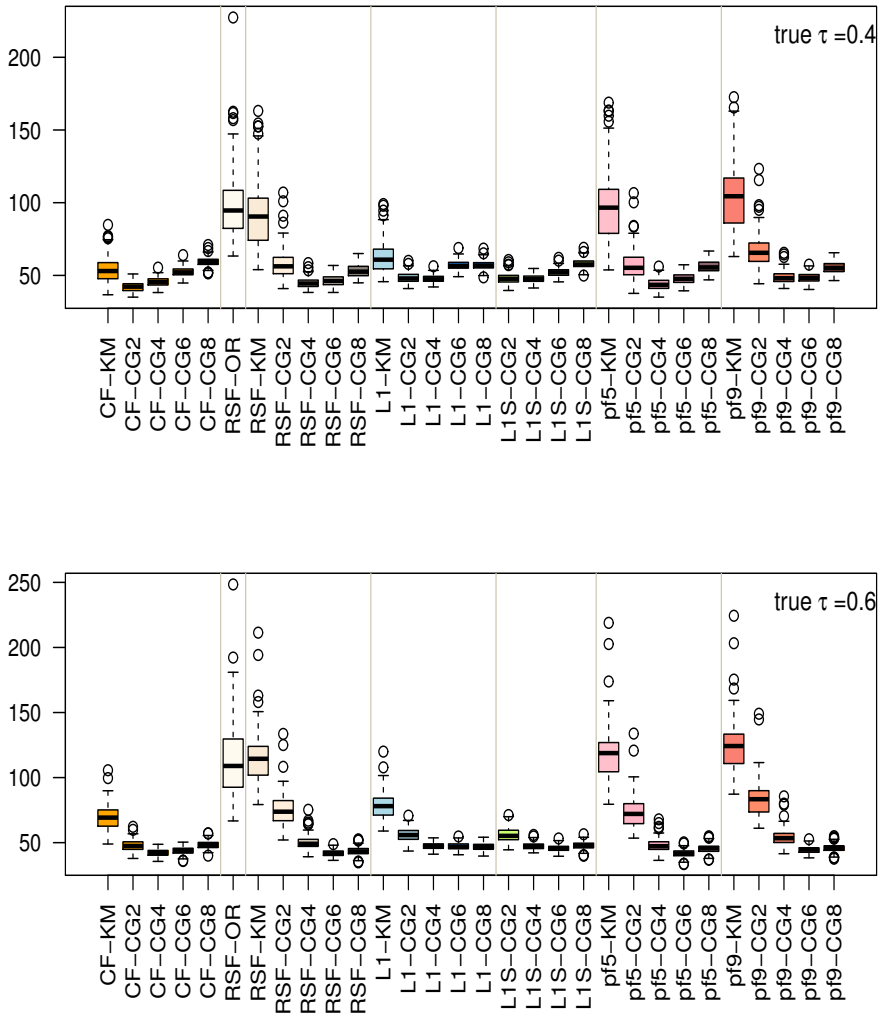

Figure 8: IAE of methods for DGP 1 at 60% censoring proportion with true τ values of 0.4 and 0.6

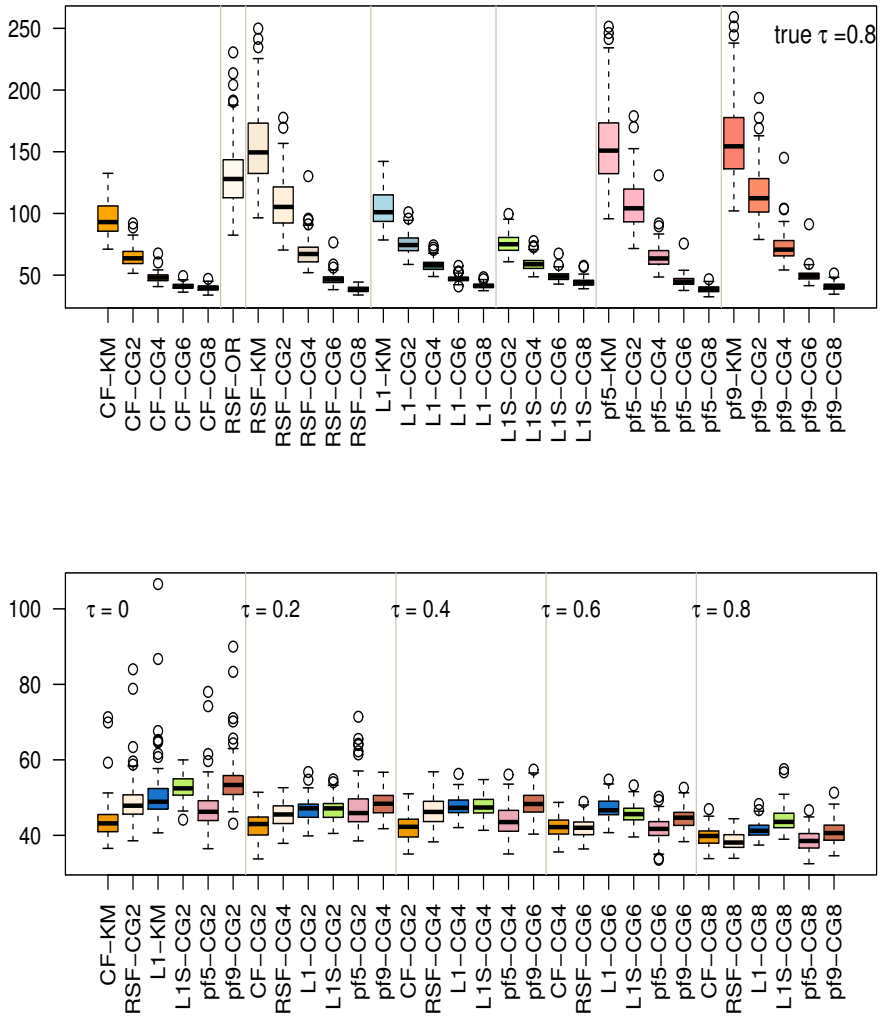

Figure 9: IAE of methods for DGP 1 at 60% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

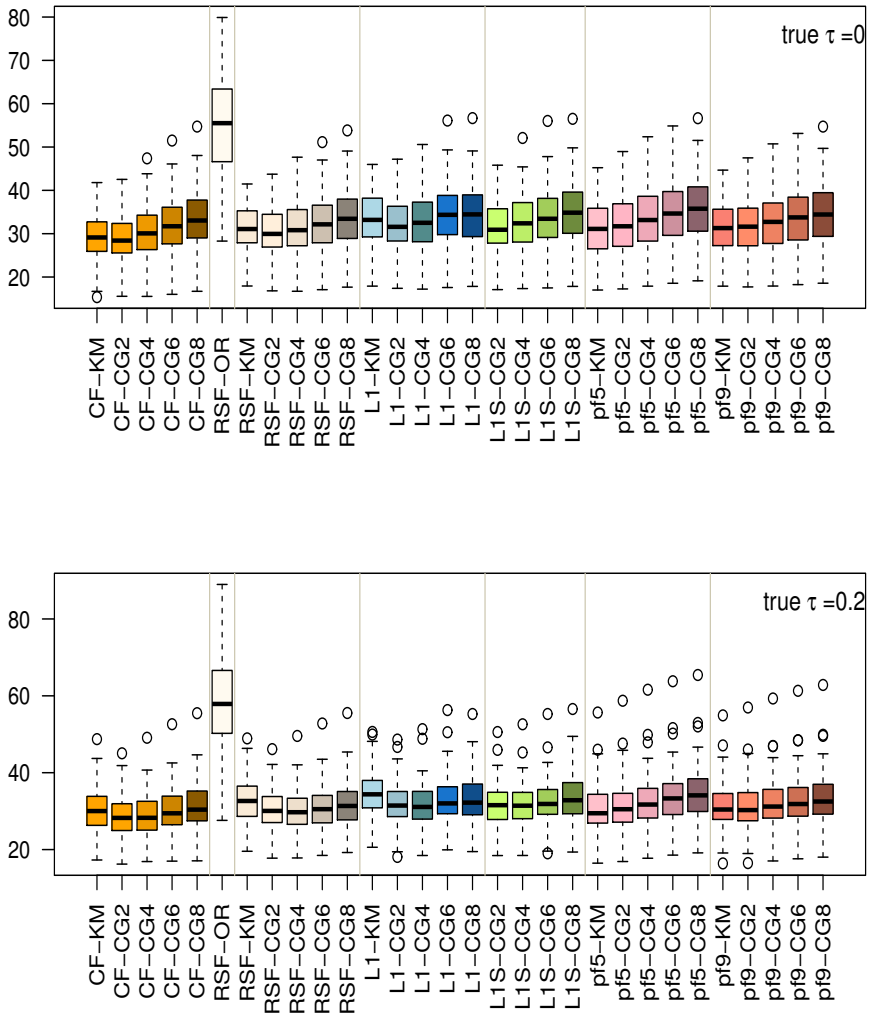

Figure 10: IAE of methods for DGP 2 at 20% censoring proportion with true τ values of 0 and 0.2

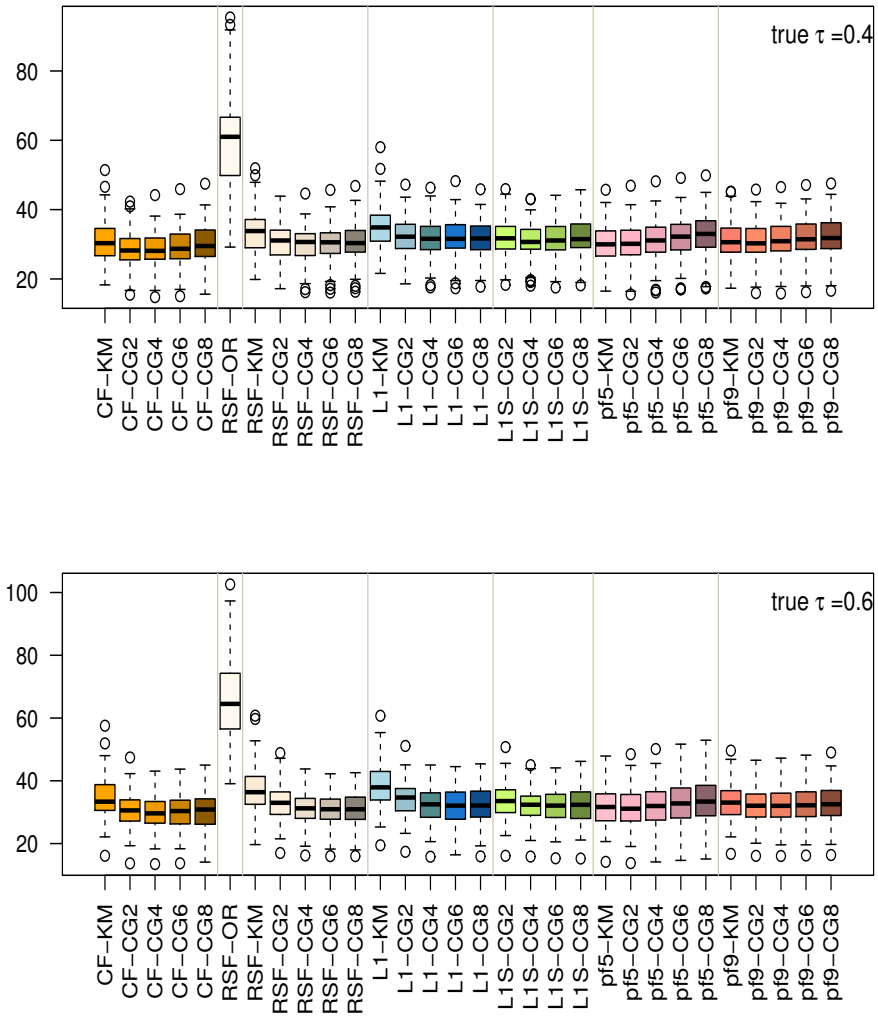

Figure 11: IAE of methods for DGP 2 at 20% censoring proportion with true τ values of 0.4 and 0.6

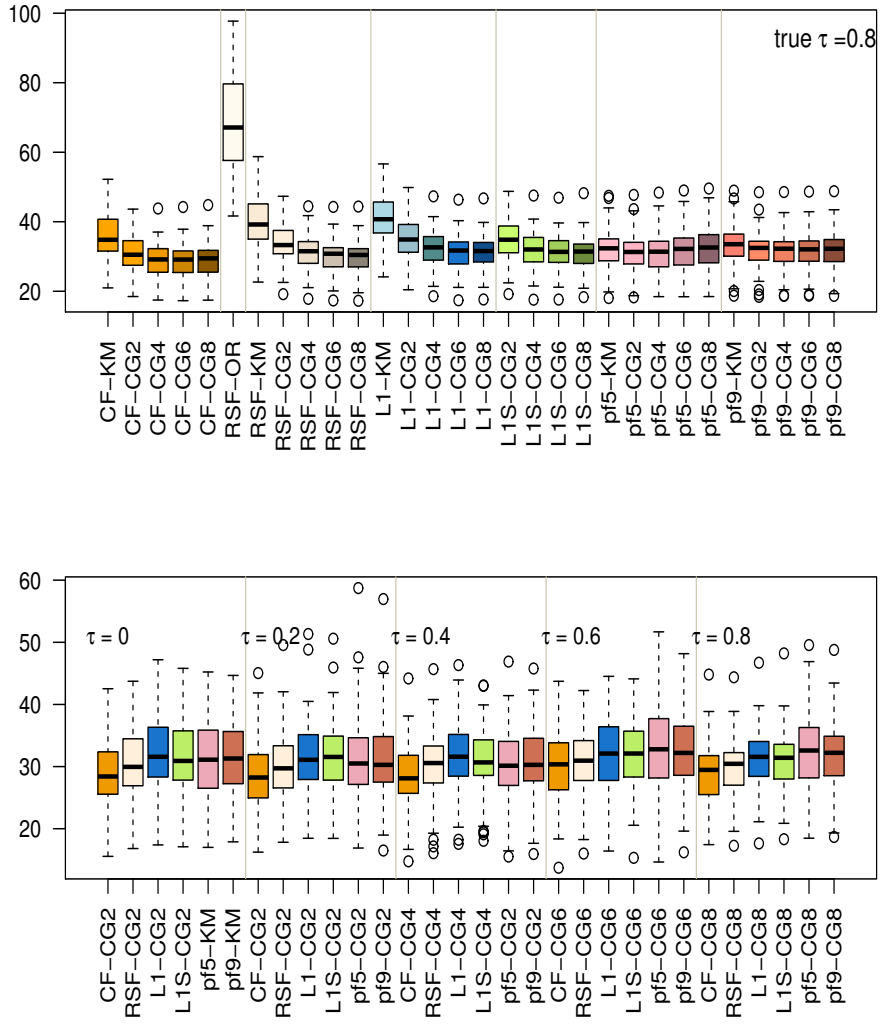

Figure 12: IAE of methods for DGP 2 at 20% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

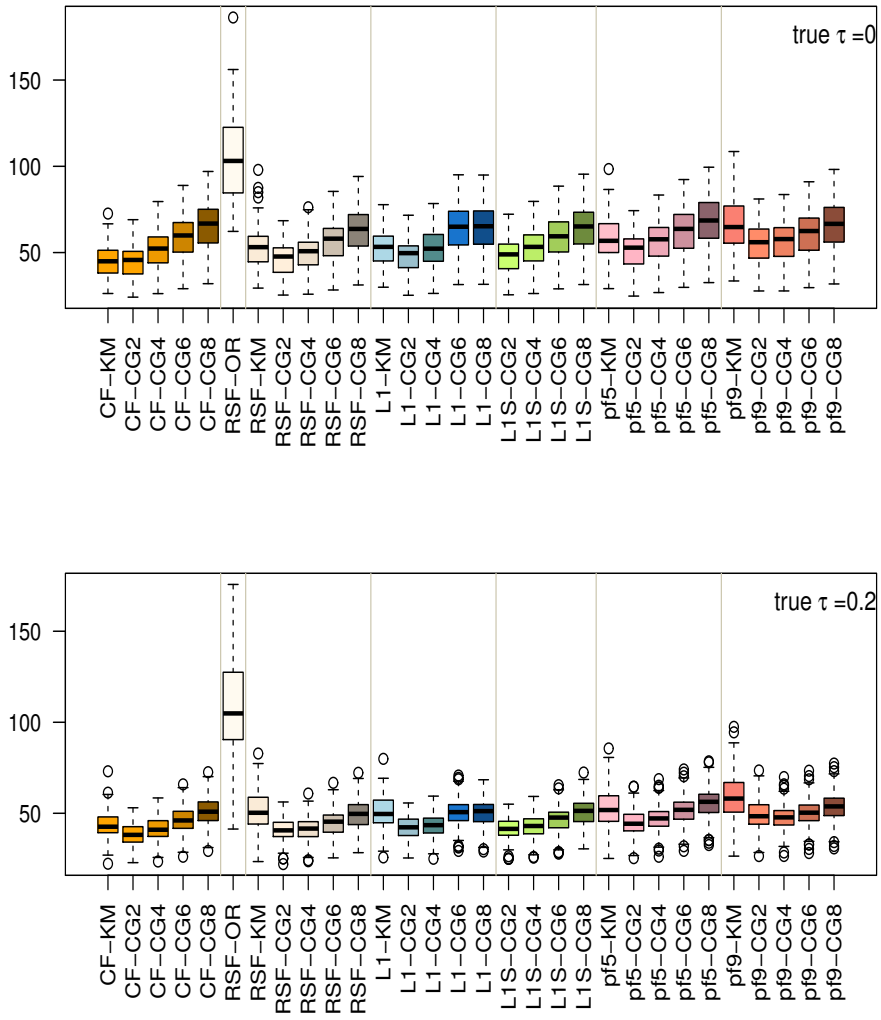

Figure 13: IAE of methods for DGP 2 at 40% censoring proportion with true τ values of 0 and 0.2

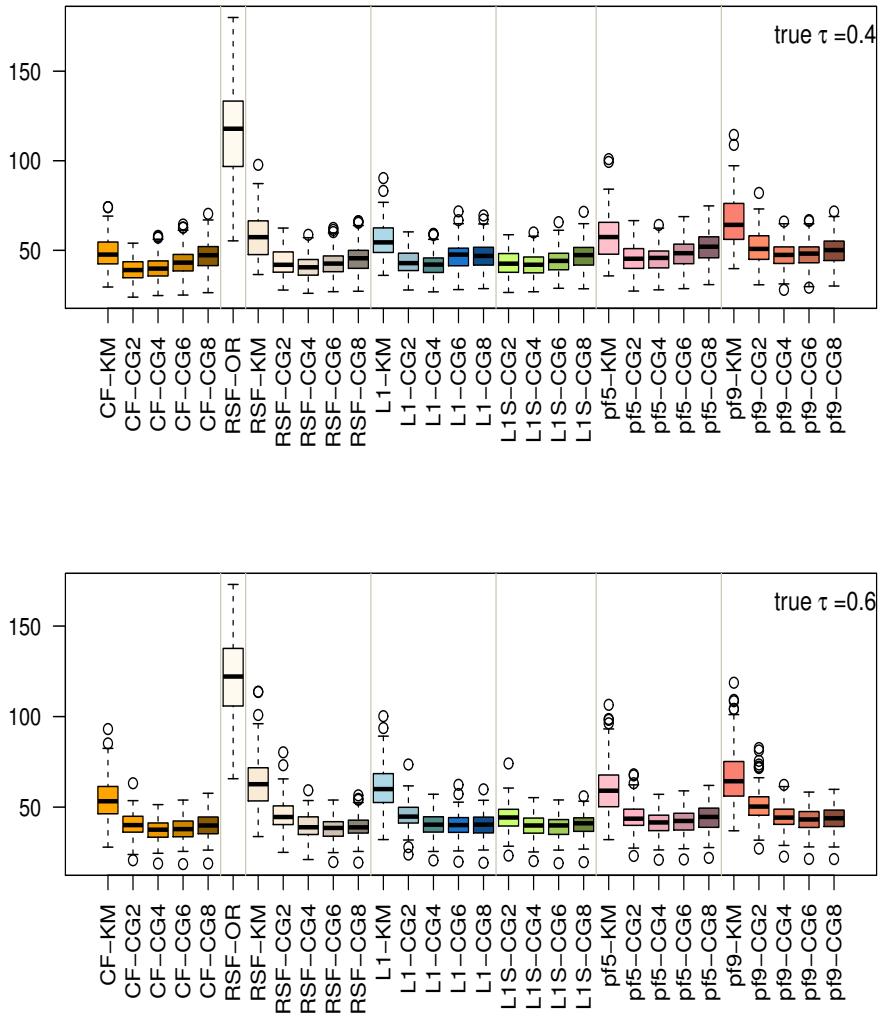

Figure 14: IAE of methods for DGP 2 at 40% censoring proportion with true τ values of 0.4 and 0.6

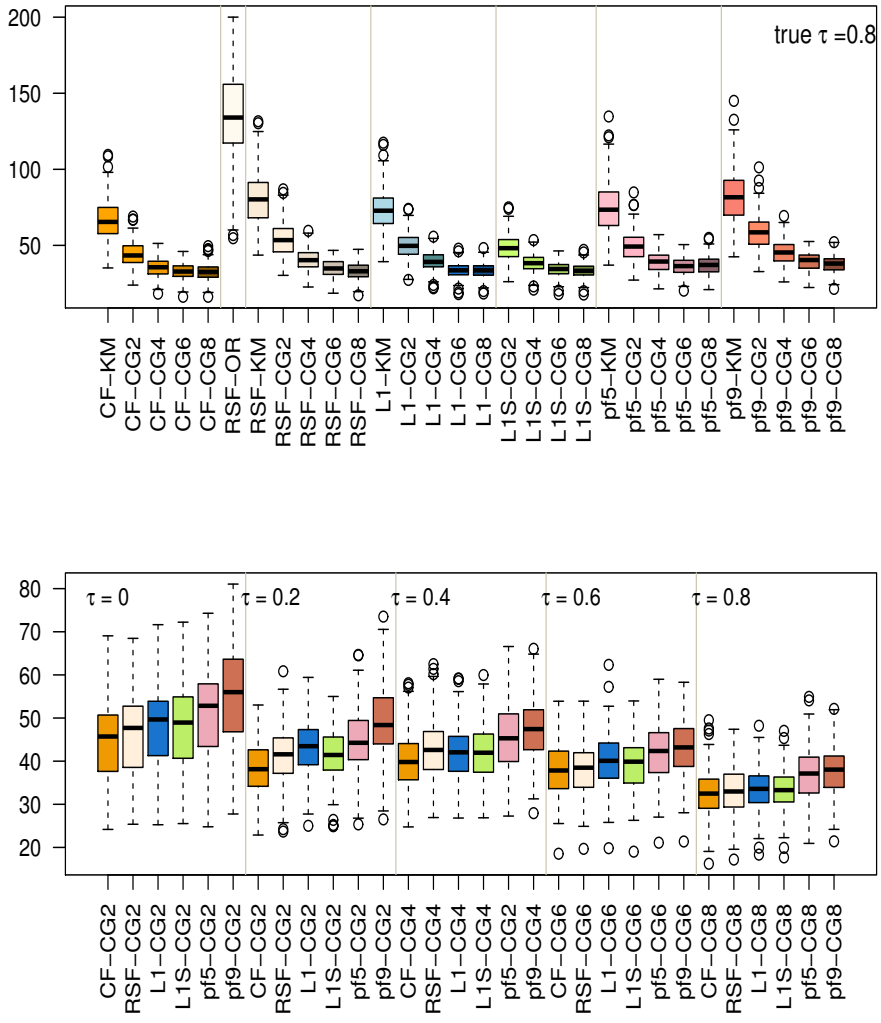

Figure 15: IAE of methods for DGP 2 at 40% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

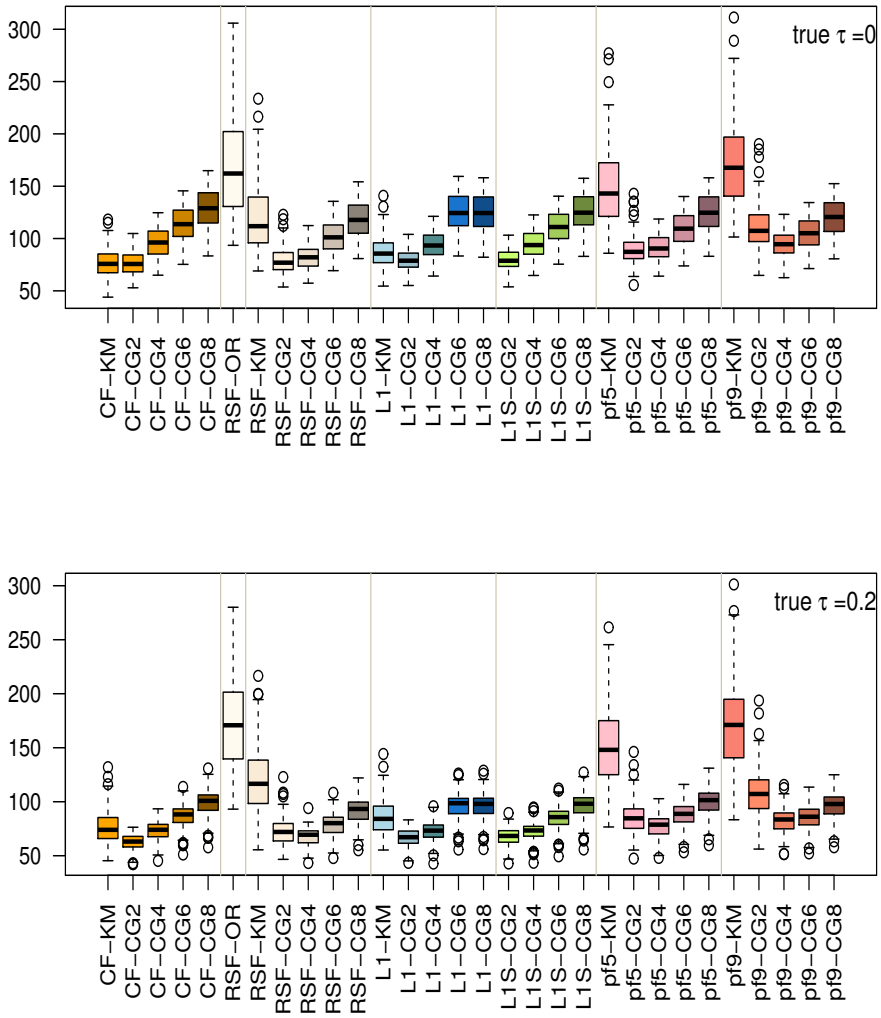

Figure 16: IAE of methods for DGP 2 at 60% censoring proportion with true τ values of 0 and 0.2

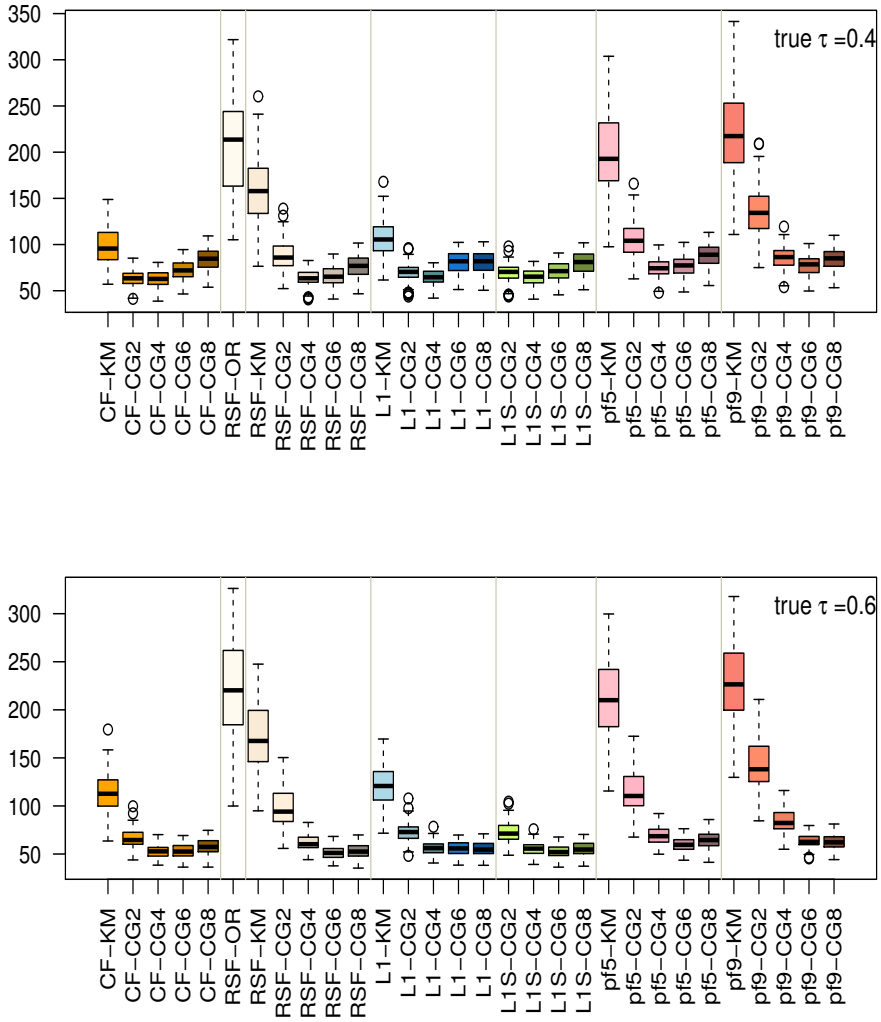

Figure 17: IAE of methods for DGP 2 at 60% censoring proportion with true τ values of 0.4 and 0.6

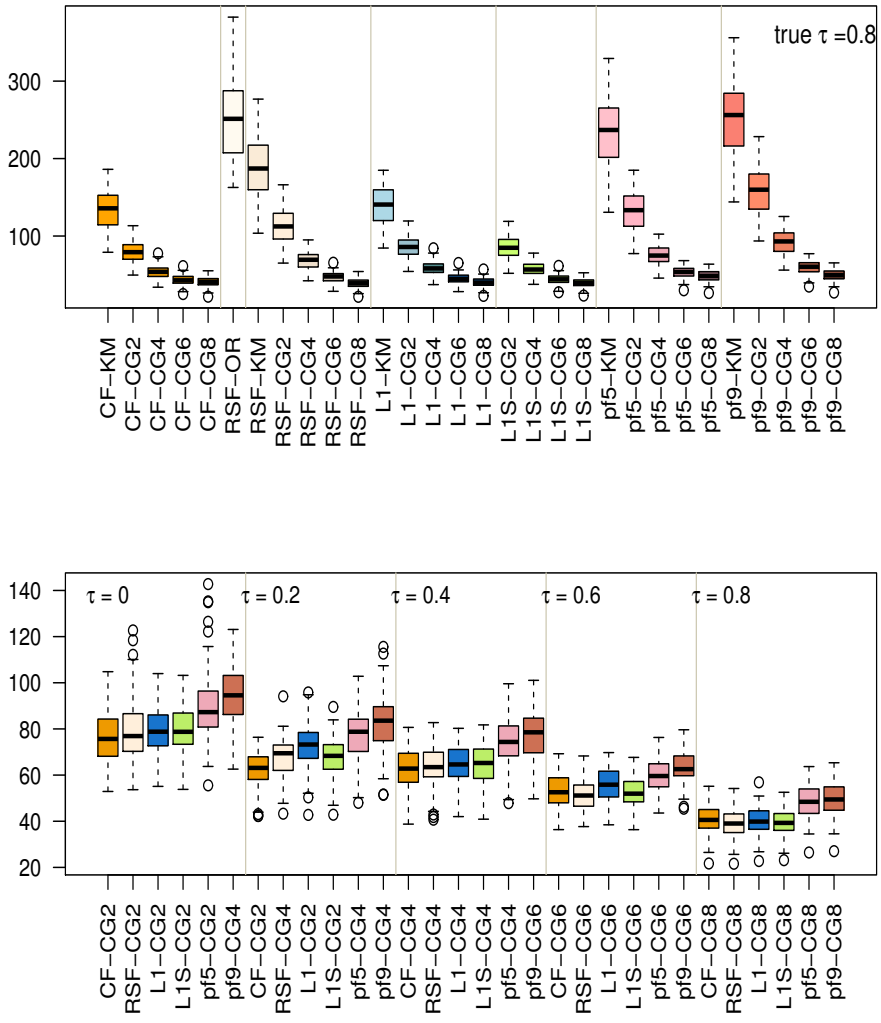

Figure 18: IAE of methods for DGP 2 at 60% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

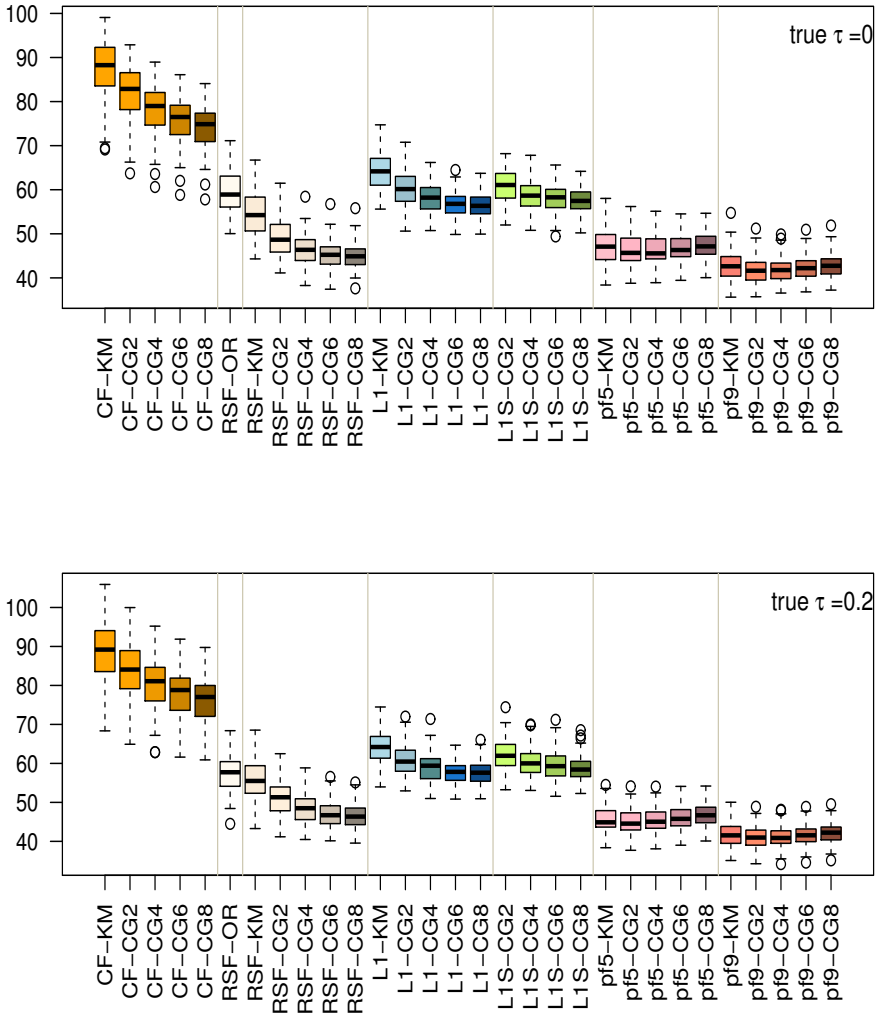

Figure 19: IAE of methods for DGP 3 at 20% censoring proportion with true τ values of 0 and 0.2

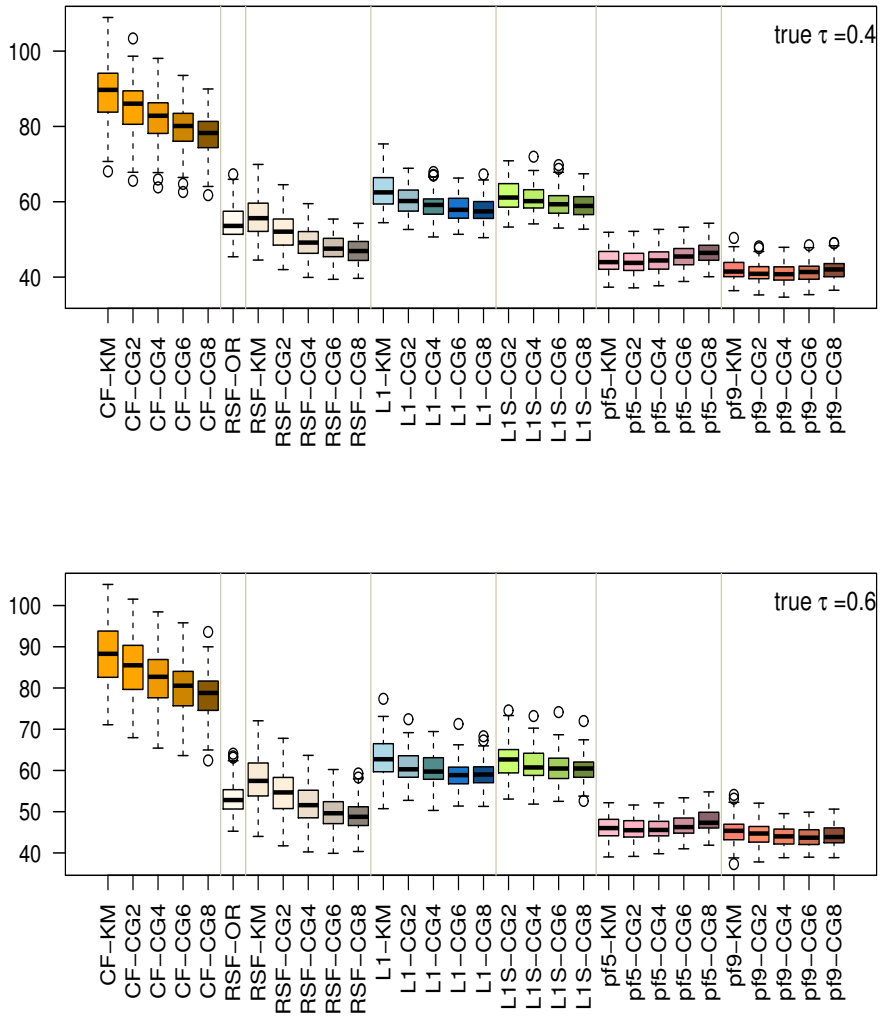

Figure 20: IAE of methods for DGP 3 at 20% censoring proportion with true τ values of 0.4 and 0.6

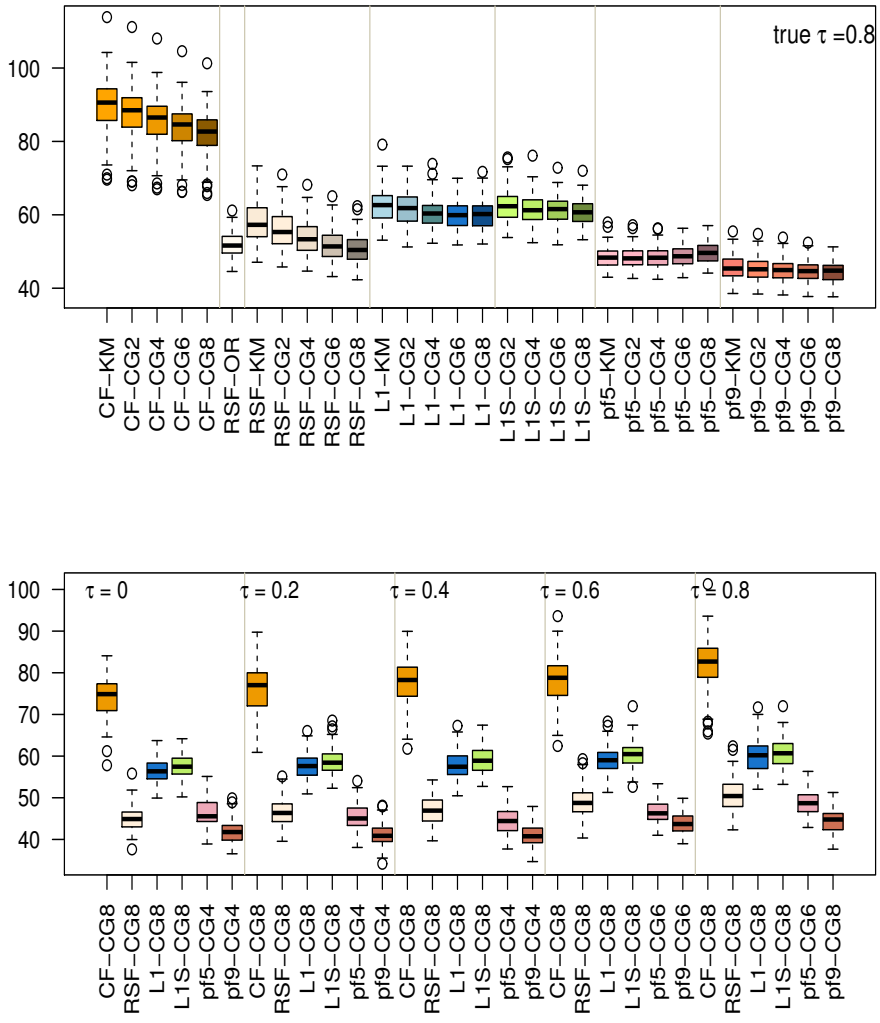

Figure 21: IAE of methods for DGP 3 at 20% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

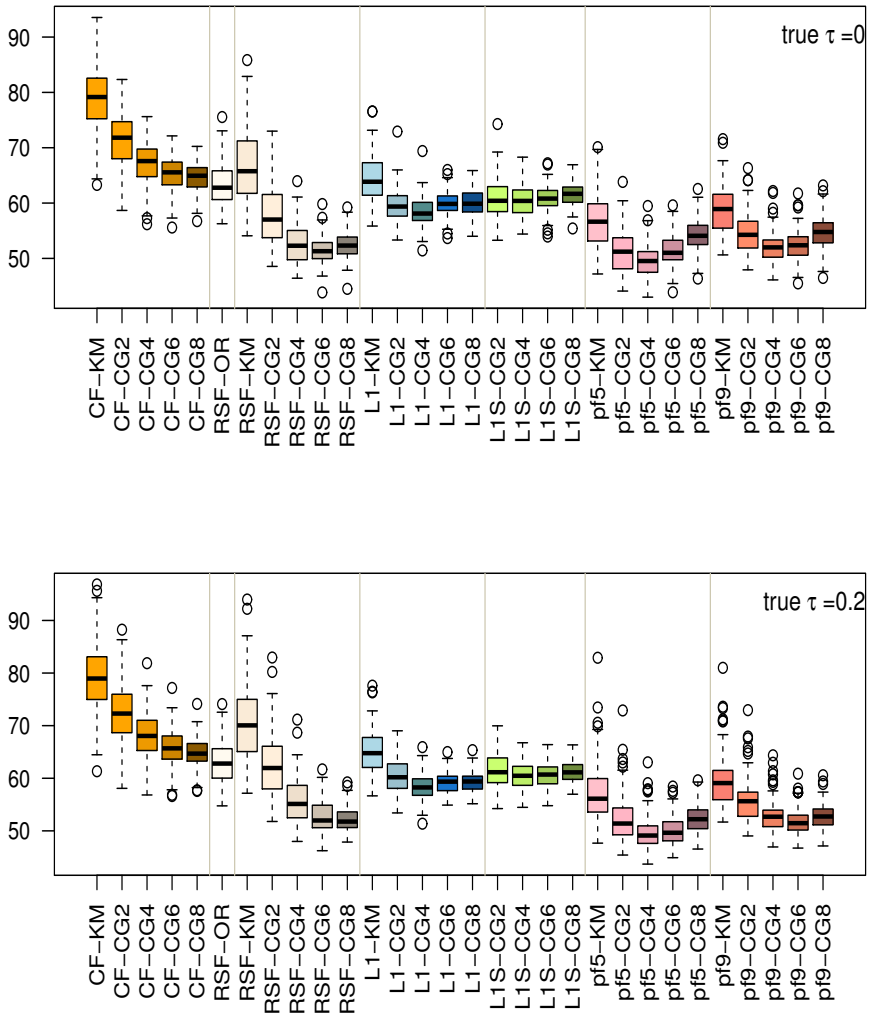

Figure 22: IAE of methods for DGP 3 at 40% censoring proportion with true τ values of 0 and 0.2

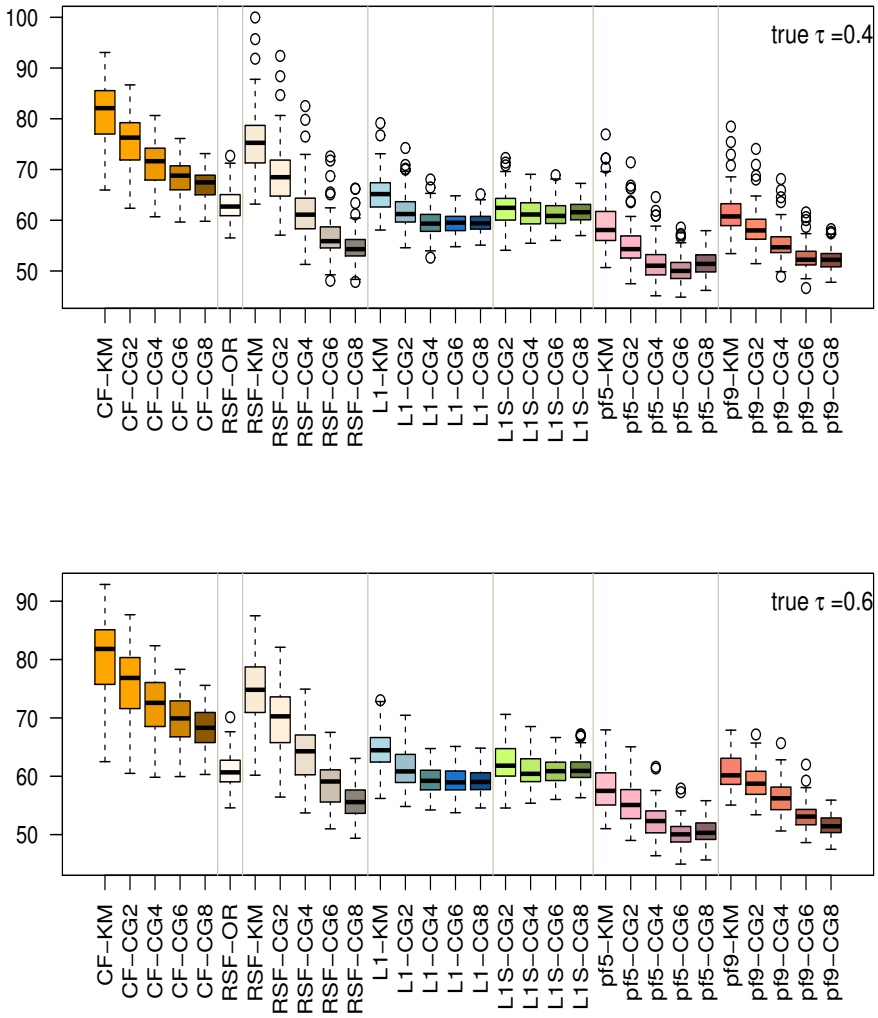

Figure 23: IAE of methods for DGP 3 at 40% censoring proportion with true τ values of 0.4 and 0.6

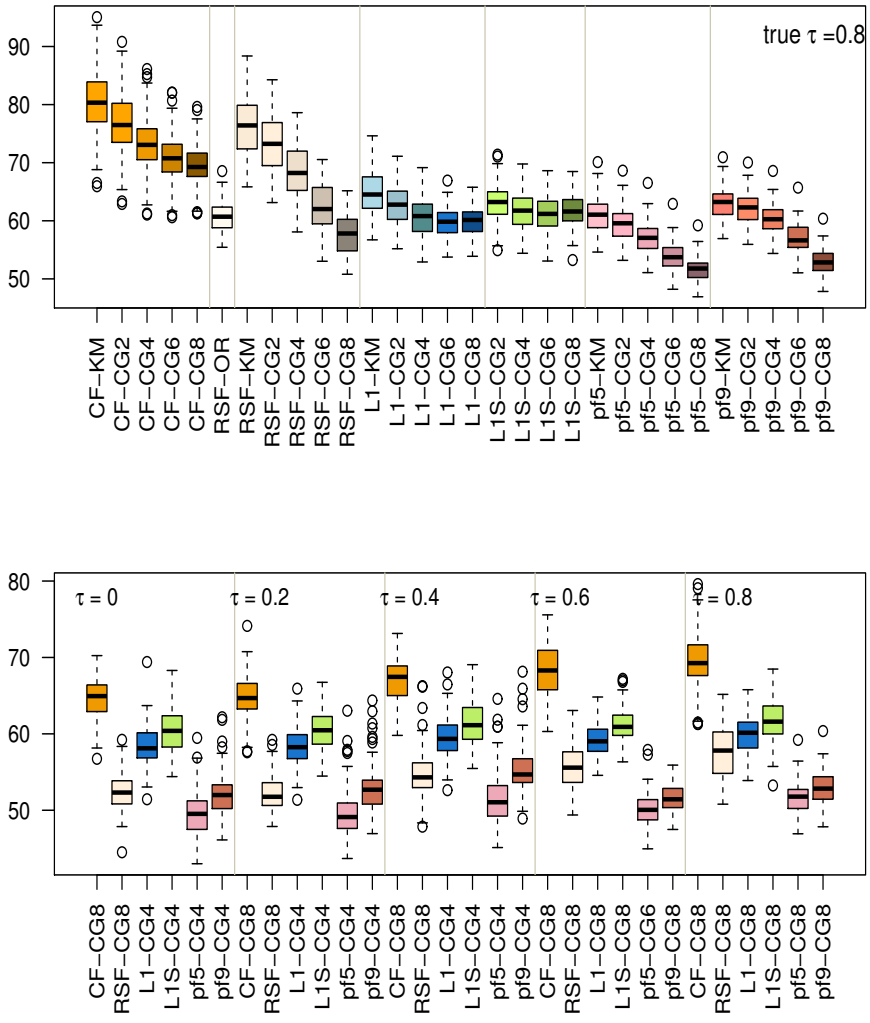

Figure 24: IAE of methods for DGP 3 at 40% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

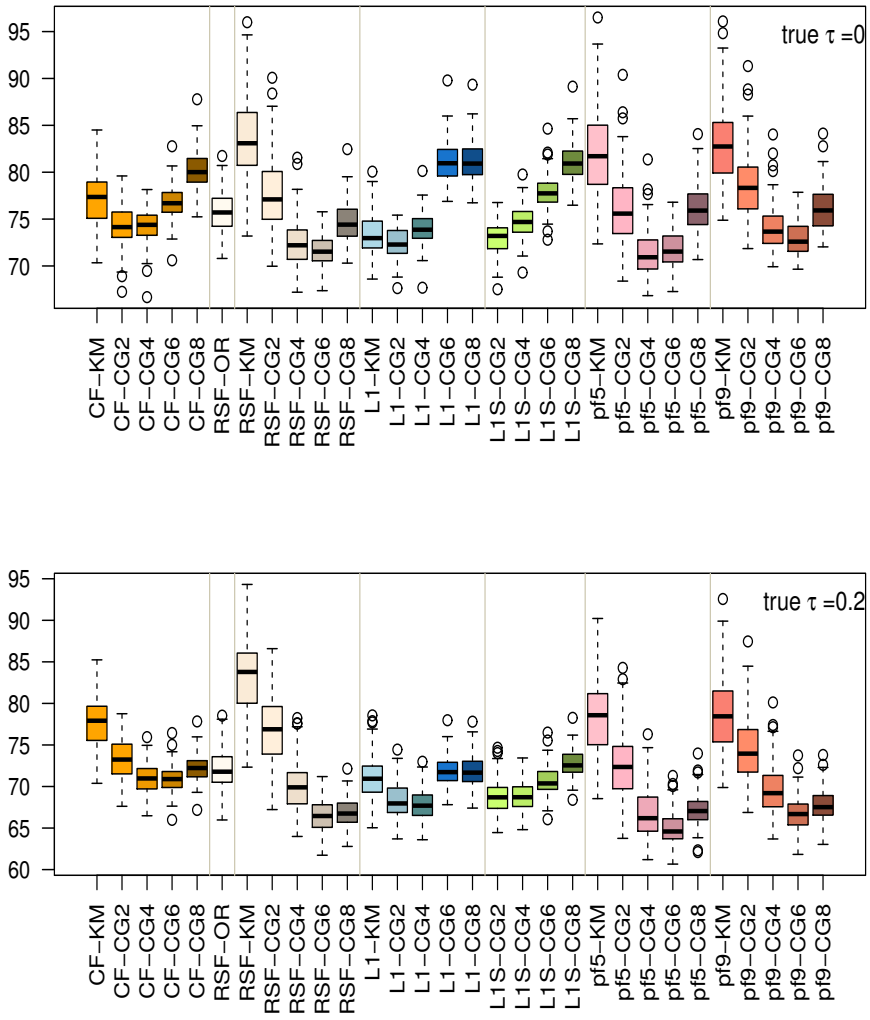

Figure 25: IAE of methods for DGP 3 at 60% censoring proportion with true τ values of 0 and 0.2

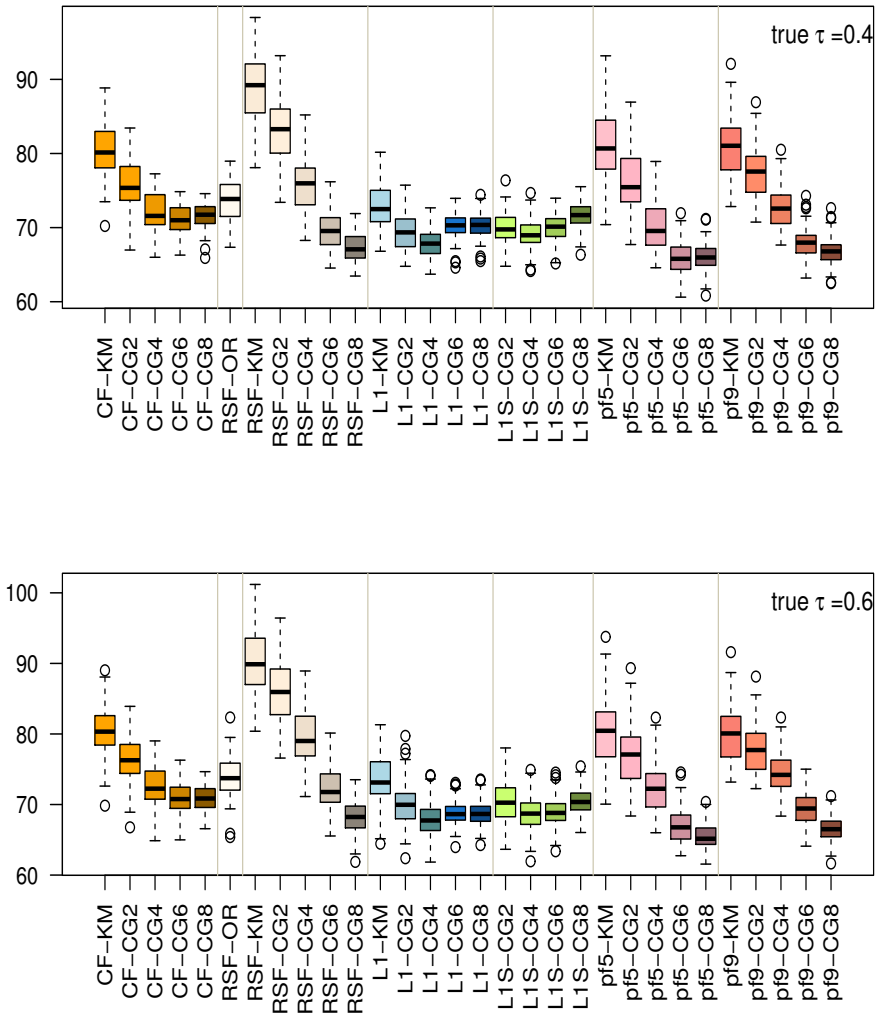

Figure 26: IAE of methods for DGP 3 at 60% censoring proportion with true τ values of 0.4 and 0.6

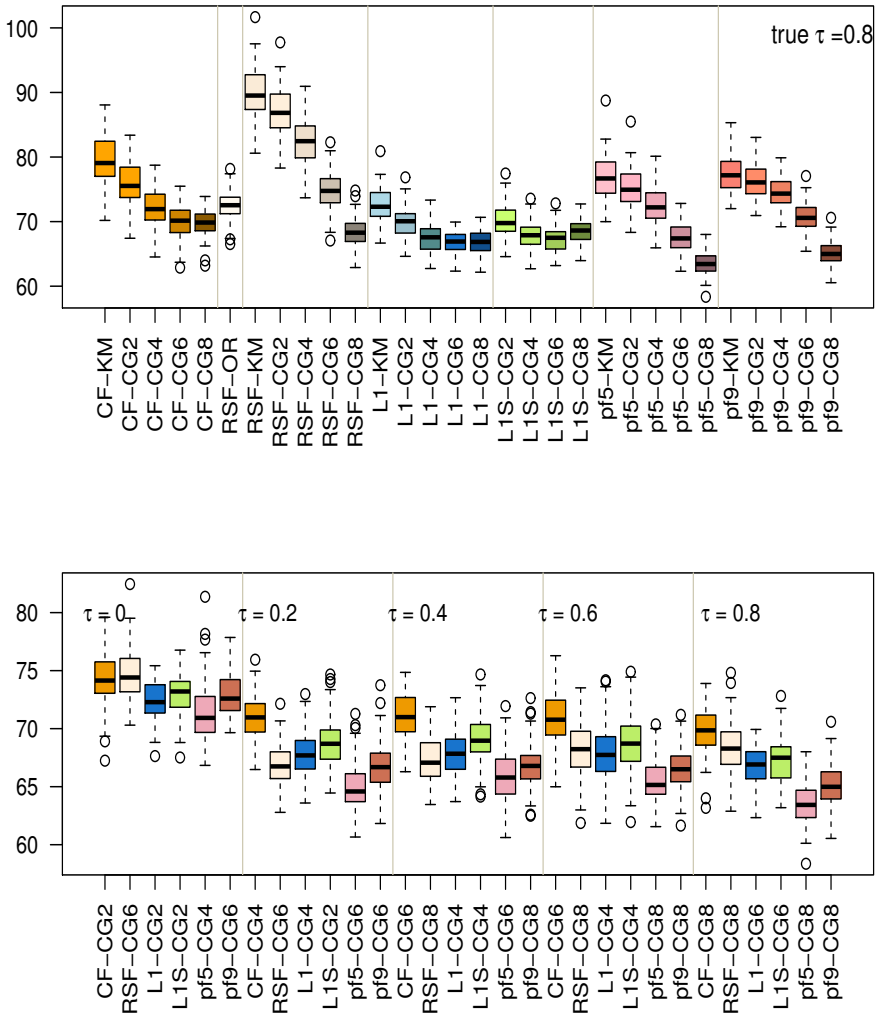

Figure 27: IAE of methods for DGP 3 at 60% censoring proportion with true τ value of 0.8 and best of each method for different values of true τ

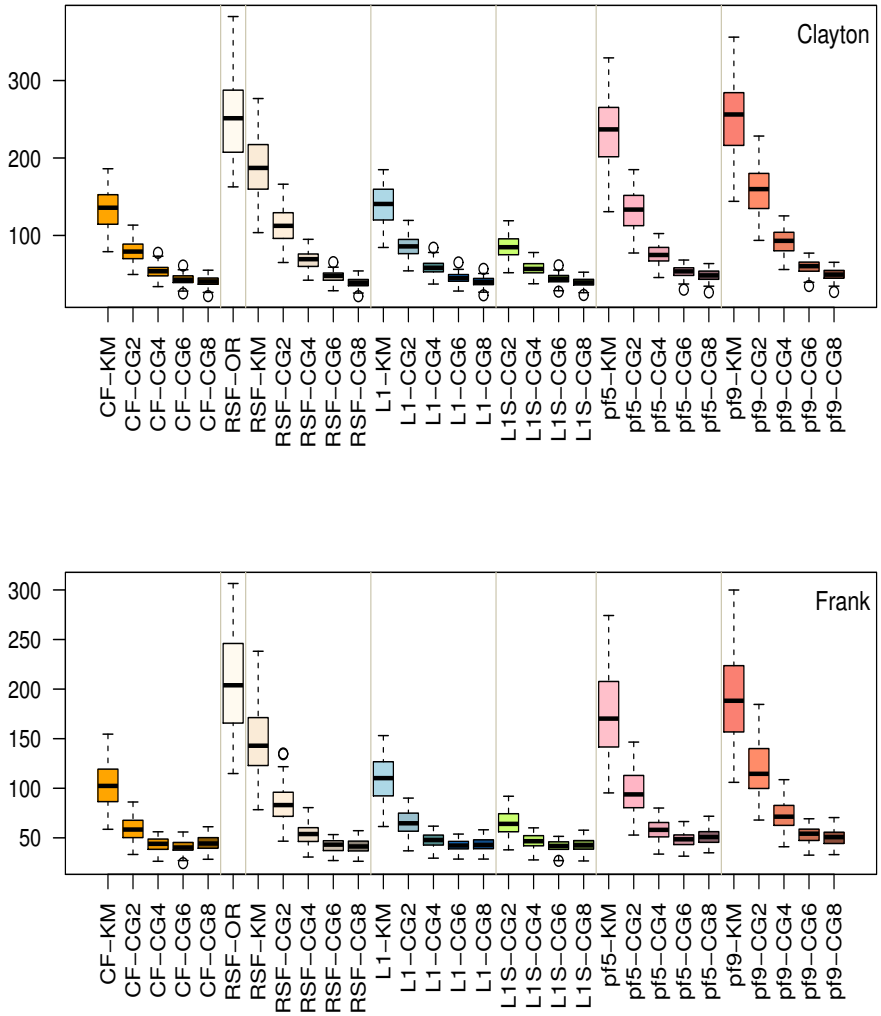

Figure 28: IAE of methods for DGP 2 produced by Clayton Copula versus Frank Copula at 60% censoring proportion with true τ values of 0.8

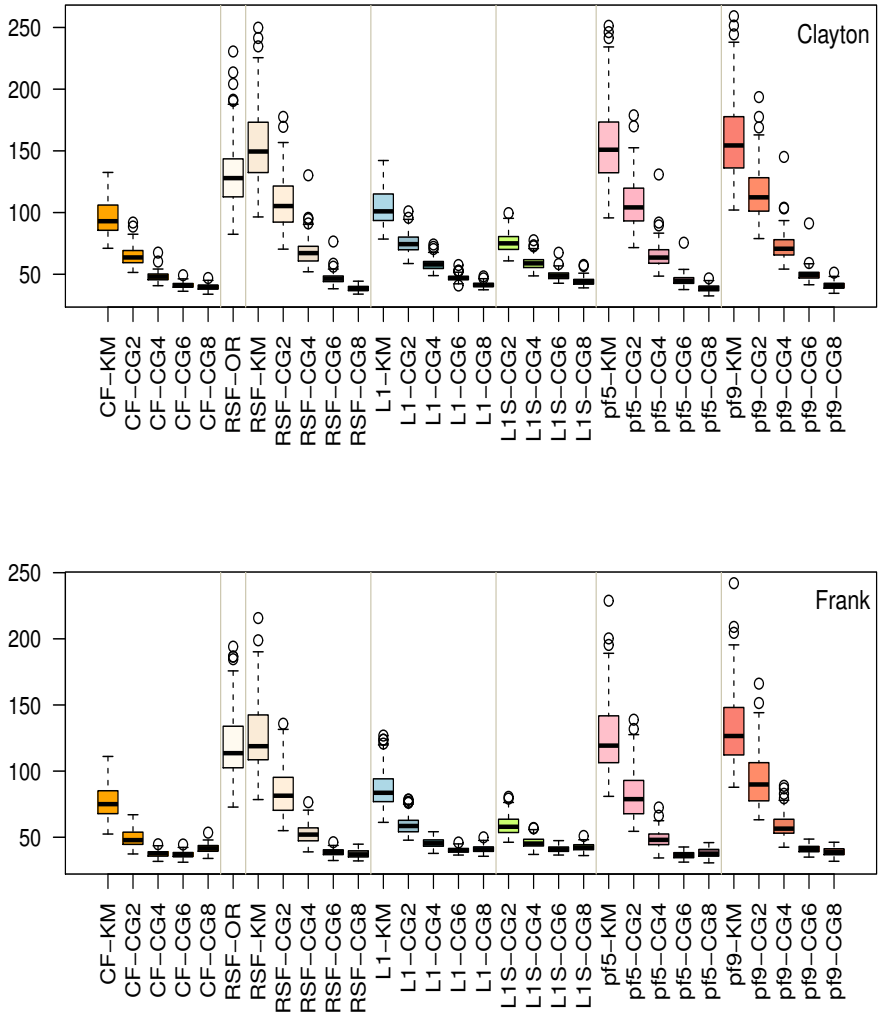

Figure 29: IAE of methods for DGP 3 produced by Clayton Copula versus Frank Copula at 60% censoring proportion with true τ values of 0.8

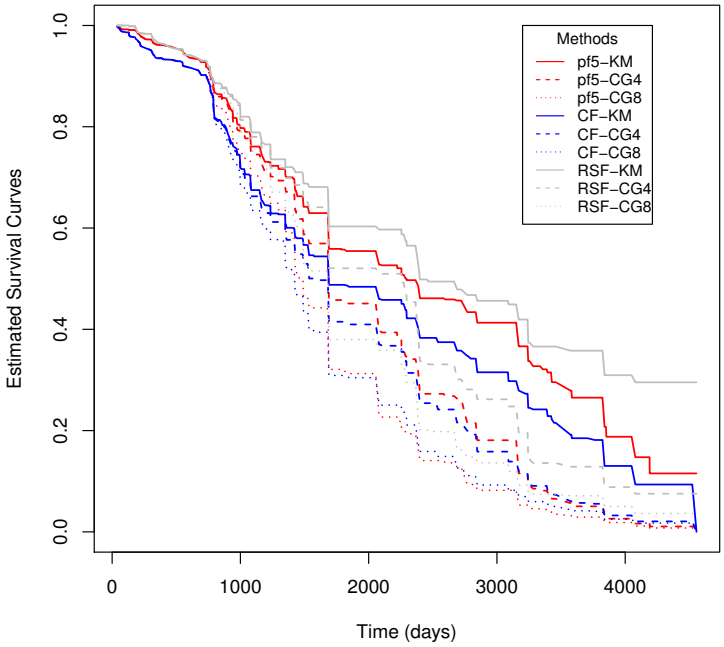

Figure 30: Estimated survival functions for a new subject

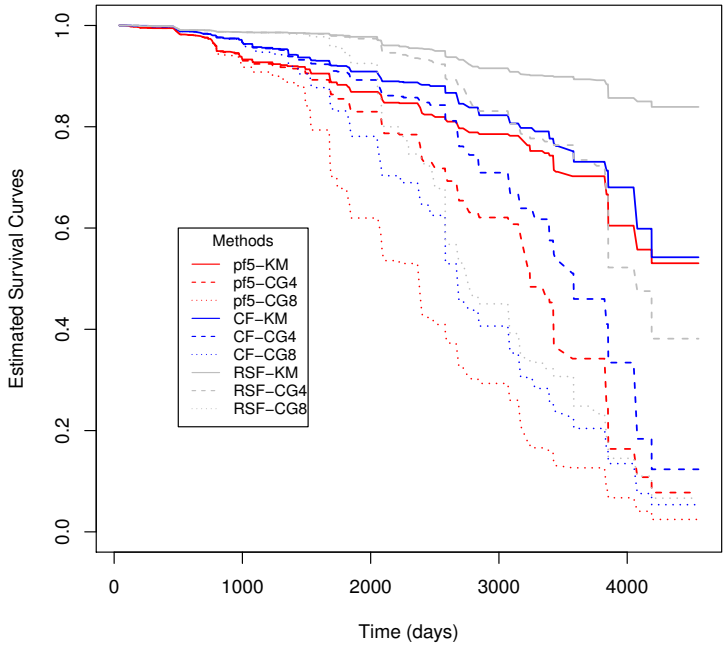

Figure 31: Estimated survival functions for a new subject

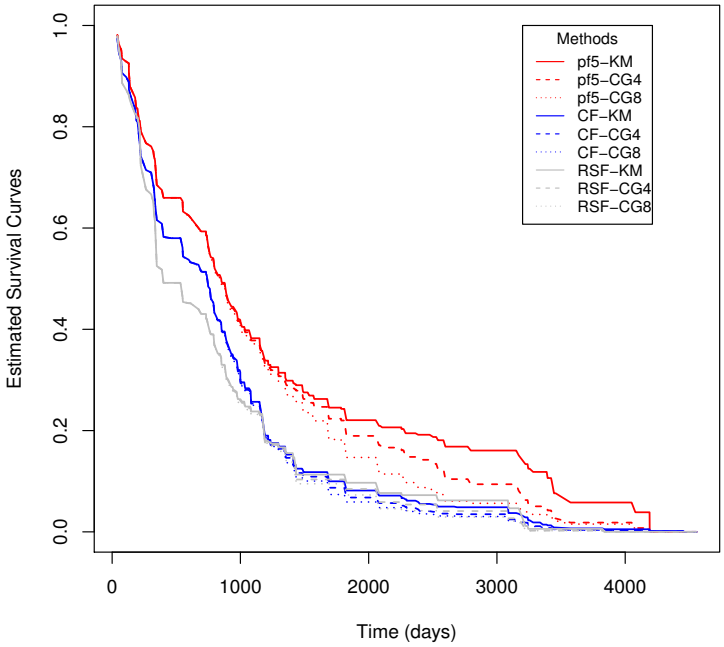

Figure 32: Estimated survival functions for a new subject

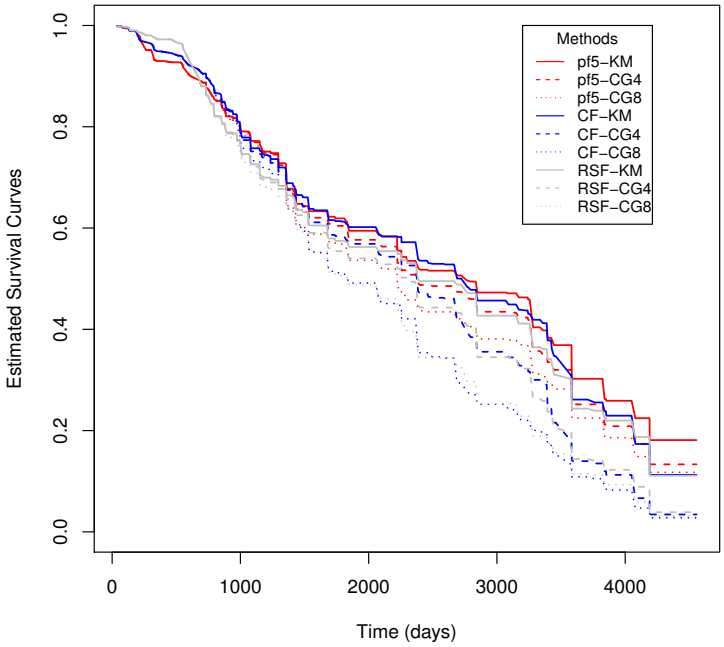

Figure 33: Estimated survival functions for a new subject

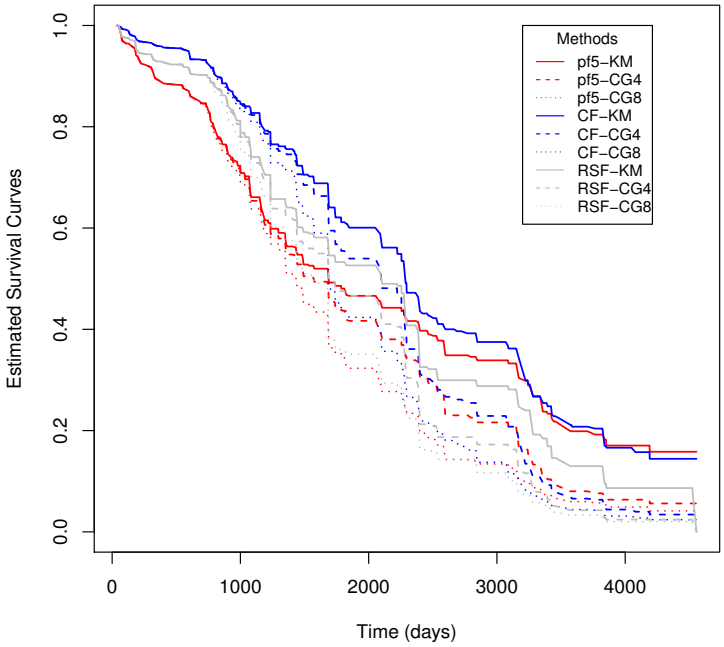

Figure 34: Estimated survival functions for a new subject

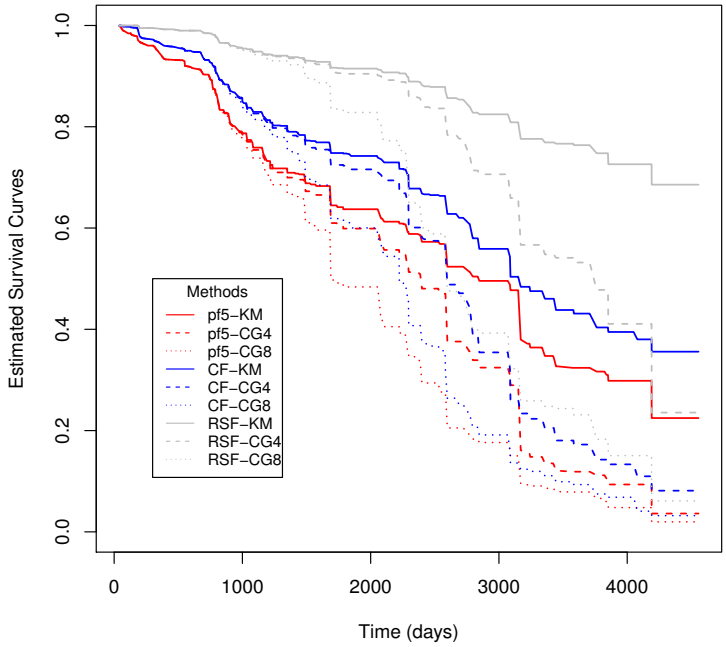

Figure 35: Estimated survival functions for a new subject

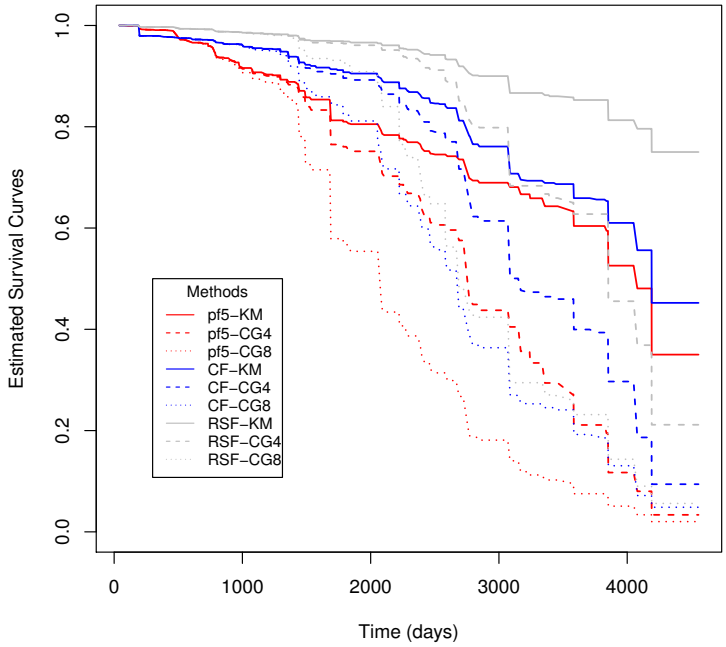

Figure 36: Estimated survival functions for a new subject

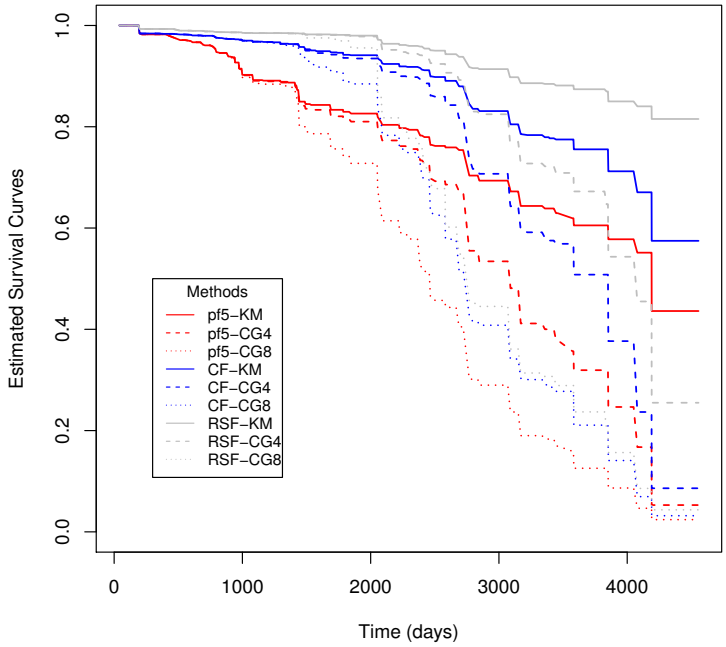

Figure 37: Estimated survival functions for a new subject

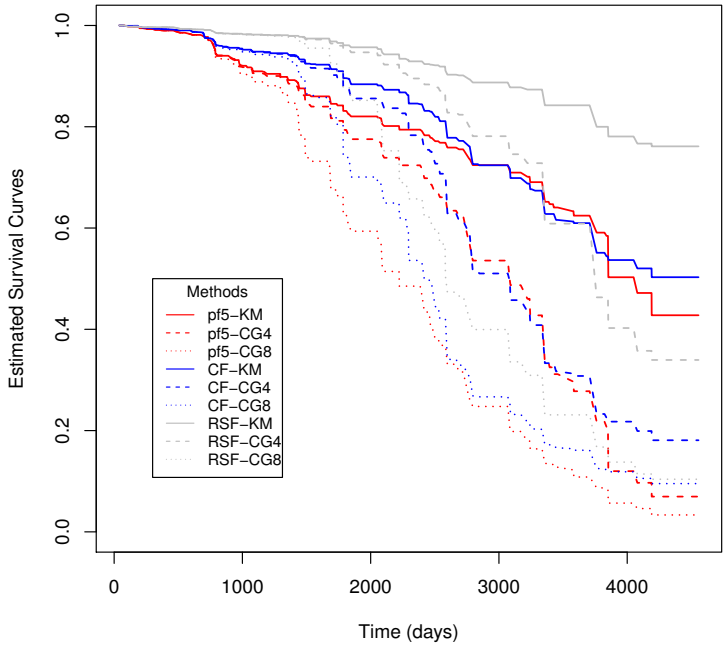

Figure 38: Estimated survival functions for a new subject