Random Forest — Reliability of Standard Error Estiamtes of Main and Low-Order Effects

Felix Kapulla

```
knitr::opts_chunk$set(fig.width=15, fig.height=8)
```

```
library(parallel)
cores <- detectCores()
clust <- makeCluster(4)
source('C:/Users/feix_/iCloudDrive/Studium Master/CQM - Thesis Internship/Thesis-VariableEffects/Rangerl
source('C:/Users/feix_/iCloudDrive/Studium Master/CQM - Thesis Internship/Thesis-VariableEffects/Reliab



parallel::clusterEvalQ(clust,
                       expr = {source('C:/Users/feix_/iCloudDrive/Studium Master/CQM - Thesis Internshi
```

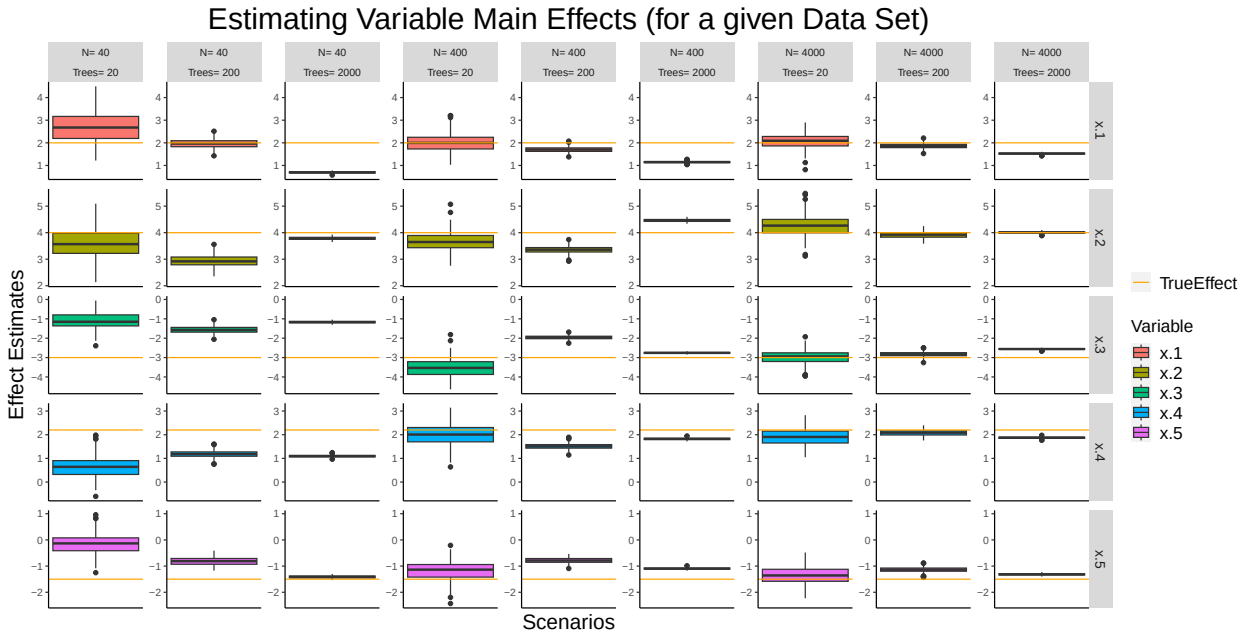
Simulation

```
### Scenarios
set.seed(123)
N <- c(40, 400, 4000) ; num.trees <- c(20, 200, 2000) ; reps <- 200; cor <- c(0)
k <- c(1); node_size <- c(1)
formulas <- c("2*x.1+4*x.2-3*x.3+2.2*x.4-1.5*x.5")
longest_latex_formula <- "2x_1+4x_2-3x_3+2.2x_4-1.5x_5"
scenarios <- data.frame(expand.grid(N, num.trees, formulas, reps,
                                    cor, k, node_size))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Reps",
                        "Correlation", "k", "Node_Size")
scenarios$k_idx <- (scenarios$k == unique(scenarios$k)[1])
scenarios[,"Formula"] <- as.character(scenarios[,"Formula"]) ### Formula became Factor
scenarios["Longest_Latex_formula"] <- longest_latex_formula
scenarios <- split(scenarios, seq(nrow(scenarios)))



system.time(result <- parLapply(cl = clust,
                                X = scenarios,
                                fun = sim))
```

```
##    user  system elapsed
##    0.36    0.42 2852.27
```

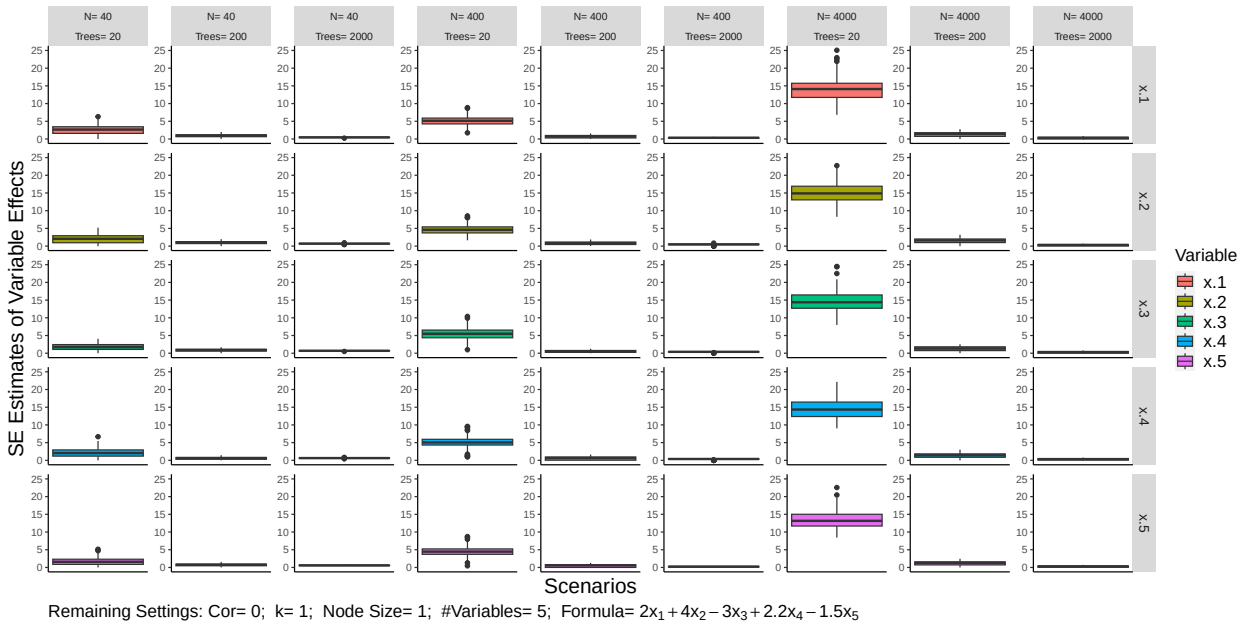
```
plot_effects(result)
```

```
## `summarise()` has grouped output by 'N', 'cor', 'k', 'num.trees', 'node_size',
## 'variable'. You can override using the `.groups` argument.
```



Remaining Settings: Cor= 0; k= 1; Node Size= 1; #Variables= 5; Formula= $2x_1 + 4x_2 - 3x_3 + 2.2x_4 - 1.5x_5$

```
plot_se_box(result)
```



Remaining Settings: Cor= 0; k= 1; Node Size= 1; #Variables= 5; Formula= $2x_1 + 4x_2 - 3x_3 + 2.2x_4 - 1.5x_5$

```
plot_se_dense(result)
```