

# Estimating Variance of Simple Defined Variable Main and Low-Order Interaction Effects

Felix Kapulla

```
knitr::opts_chunk$set(fig.width=22, fig.height=9)

library(Matrix)
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(ranger)
library(MixMatrix)
library(mvtnorm)
library(stringr)
library(parallel)

cores <- detectCores()
clust <- makeCluster(4)

source('C:/Users/feix_/iCloudDrive/Studium Master/CQM - Thesis Internship/Thesis-VariableEffects/Baseli

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

## Simulation

```
n <- c(40, 400, 4000) ; num.trees <- 2000 ; repeats <- 200; cor <- c(0, 0.8)
k <- c(0.2, 1); node_size <- c(1); pdp <- F; ale <- F

formulas <- c("2*x.1+4*x.2-3*x.3+2.2*x.4")
#parallel::clusterExport(cl = clust, varlist = 'formulas')
scenarios <- data.frame(expand.grid(n, num.trees, formulas, repeats,
                                   cor, k, node_size, pdp, ale))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats",
                       "Correlation", "k", "Node_Size", "pdp", "ale")
scenarios$k_idx <- (scenarios$k == unique(scenarios$k)[1])
scenarios[, "Formula"] <- as.character(scenarios[, "Formula"]) ### Formula became Factor
scenarios <- split(scenarios, seq(nrow(scenarios)))
#Run Simulation
system.time(result <- parLapply(cl = clust,
                                X = scenarios,
                                fun = sim_multi))
```

```
##      user      system elapsed
##      0.02       0.27 11315.59
```

```
if (!pdp | !ale) {
  print_results(result)
}
```

```
## Setting 1: N = 40 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##      Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      1.244046 3.135649 -2.363102 1.393687
## Mean(s) of simulated LM Variable Effect(s):
##      2.024413 3.995613 -3.007246 2.19252
## True Variable Effect(s):
##      2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
##      1.400423 2.582836 1.937675 1.468733 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      1.450204 2.712493 2.289122 1.599211 .
## Number of Smaller Nulls:
##      0 0 0 0
##
## Setting 2: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##      Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      1.77624 4.191807 -3.159551 2.07336
## Mean(s) of simulated LM Variable Effect(s):
##      1.99696 4.004295 -2.994332 2.202076
## True Variable Effect(s):
##      2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
##      0.8170386 1.681896 1.398941 0.9746578 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.8994434 1.776224 1.516968 1.020987 .
## Number of Smaller Nulls:
##      21 2 5 23
##
## Setting 3: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##      Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      2.115773 4.413908 -3.228033 2.382982
## Mean(s) of simulated LM Variable Effect(s):
##      2.001072 3.999964 -3.000873 2.200079
## True Variable Effect(s):
##      2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
##      0.6796585 1.064733 0.8940101 0.7746687 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      1.014772 1.22639 1.116316 0.9986079 .
## Number of Smaller Nulls:
##      37 34 45 44
##
## Setting 4: N = 40 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##      Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
```

```

## Mean(s) of simulated RF Variable Effect(s):
## 1.431335 2.745878 -0.4384682 1.490392
## Mean(s) of simulated LM Variable Effect(s):
## 1.986593 4.008304 -2.988236 2.165546
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 1.561905 1.784588 0.6438469 1.402807 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 1.596612 2.181109 0.7981635 1.598454 .
## Number of Smaller Nulls:
## 0 0 1 0
##
## Setting 5: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 1.500982 3.97431 -1.93095 1.911254
## Mean(s) of simulated LM Variable Effect(s):
## 2.012684 4.006641 -3.012013 2.191657
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.9905318 1.361753 0.6890726 1.160768 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 1.076303 1.621776 0.9013898 1.184026 .
## Number of Smaller Nulls:
## 2 0 3 2
##
## Setting 6: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 2.056959 3.966914 -2.859657 2.240523
## Mean(s) of simulated LM Variable Effect(s):
## 2.001127 3.999582 -2.999414 2.198854
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.9274358 0.997255 0.9104514 0.8951138 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.9577475 1.204441 1.007023 1.053774 .
## Number of Smaller Nulls:
## 39 23 31 28
##
## Setting 7: N = 40 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 1.087082 2.943865 -1.978102 1.300898
## Mean(s) of simulated LM Variable Effect(s):
## 1.972416 4.010048 -3.014445 2.192747
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.5701989 0.724896 0.6613681 0.5905404 .
## Mean of Standard Errors Estimates of Variable Effects (RF):

```

```

## 0.6174196 0.8601907 0.7590344 0.6624634 .
## Number of Smaller Nulls:
## 0 0 0 0
##
## Setting 8: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 1.620792 3.838898 -2.80765 1.847068
## Mean(s) of simulated LM Variable Effect(s):
## 2.002008 3.999918 -3.003408 2.203761
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.2926803 0.3924835 0.3711839 0.3410713 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.3312817 0.4463657 0.429213 0.3644785 .
## Number of Smaller Nulls:
## 10 3 6 4
##
## Setting 9: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 1.903866 3.949649 -2.930254 2.108479
## Mean(s) of simulated LM Variable Effect(s):
## 2.000503 4.001484 -3.001625 2.199167
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.2303815 0.2280375 0.2287054 0.2224156 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.265117 0.2715618 0.2913043 0.2563928 .
## Number of Smaller Nulls:
## 33 42 32 38
##
## Setting 10: N = 40 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 1.335443 2.41507 -0.2829194 1.391017
## Mean(s) of simulated LM Variable Effect(s):
## 1.997534 4.002252 -3.016933 2.200675
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.5201443 0.5399277 0.326208 0.5228899 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.5649971 0.6215424 0.3588624 0.5700134 .
## Number of Smaller Nulls:
## 0 0 0 0
##
## Setting 11: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 1.402701 3.354758 -1.427531 1.61378
## Mean(s) of simulated LM Variable Effect(s):

```

```
## 1.989217 3.988183 -2.996645 2.214799
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.3373745 0.3855339 0.2713698 0.2866192 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.3461795 0.4280332 0.2856964 0.366536 .
## Number of Smaller Nulls:
## 2 1 5 3
##
## Setting 12: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
## Formula = 2*x.1+4*x.2-3*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## 1.720693 3.85578 -2.315296 1.937454
## Mean(s) of simulated LM Variable Effect(s):
## 2.002184 4.000801 -3.000477 2.19719
## True Variable Effect(s):
## 2 4 -3 2.2
## Standard Error of simulated Variable Effects (RF):
## 0.2185767 0.2740843 0.2185751 0.2581528 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.2423496 0.2712983 0.227902 0.2567276 .
## Number of Smaller Nulls:
## 35 30 32 28
```

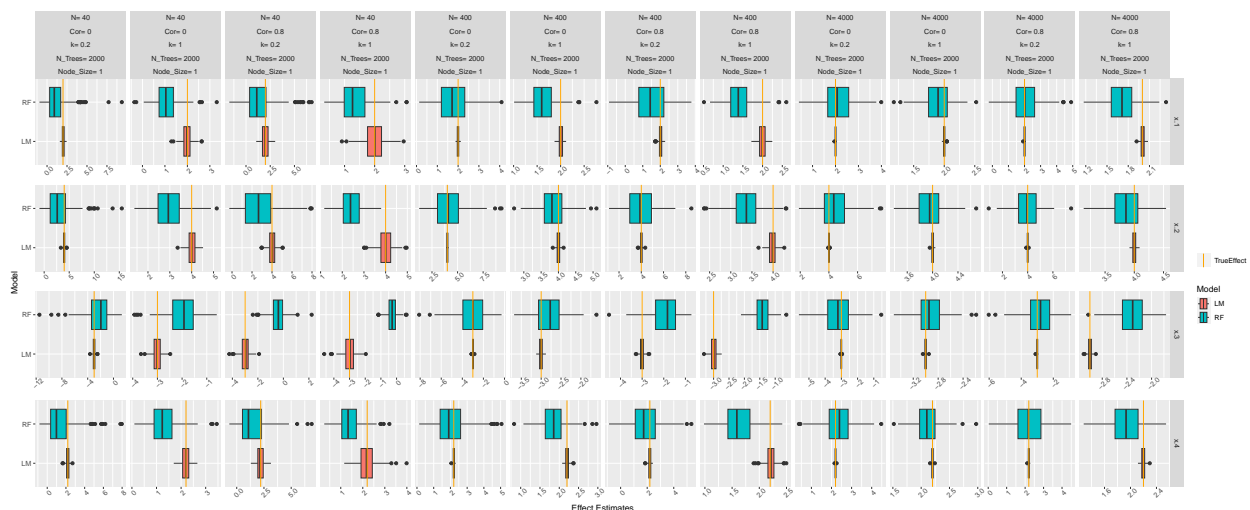
```
effect_plots <- plot_effects(result)
```

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

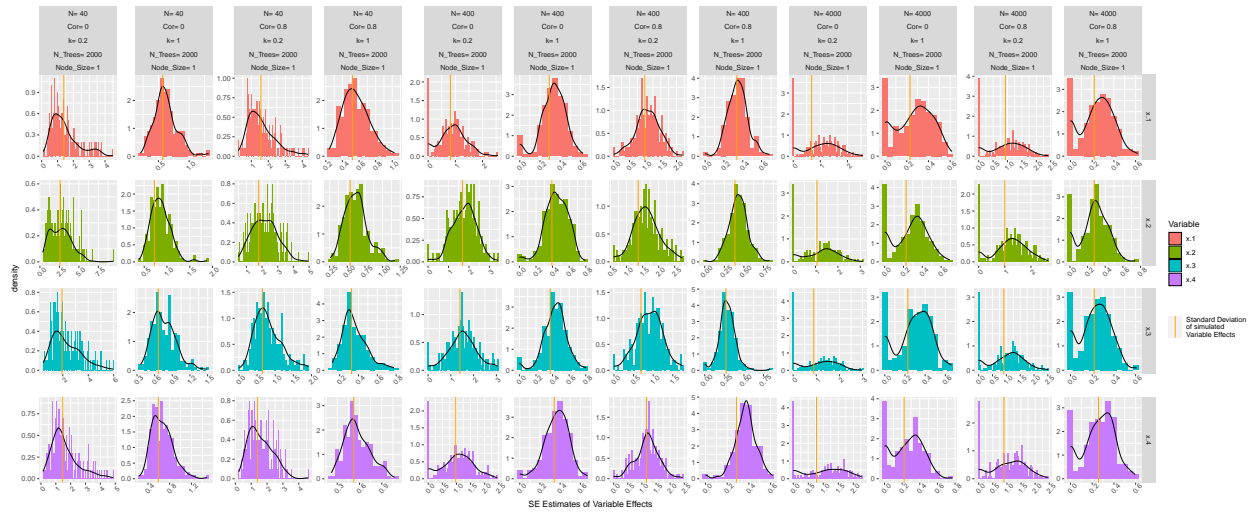
```
se_plot <- plot_se(result)
```

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

```
effect_plots
```



## se\_plot



```
if (pdp | ale) {
  plot_marginal(result)
}
```

```

n <- c(40, 400, 4000) ; num.trees <- 2000 ; repeats <- 200; cor <- c(0, 0.8)
k <- c(0.2, 1); node_size <- c(1); pdp <- F; ale <- F

formulas <- c("-0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)")
#parallel::clusterExport(cl = clust, varlist = 'formulas')
scenarios <- data.frame(expand.grid(n, num.trees, formulas, repeats,
                                   cor, k, node_size, pdp, ale))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats",
                       "Correlation", "k", "Node_Size", "pdp", "ale")
scenarios$k_idx <- (scenarios$k == unique(scenarios$k)[1])
scenarios[, "Formula"] <- as.character(scenarios[, "Formula"]) ### Formula became Factor
scenarios <- split(scenarios, seq(nrow(scenarios)))
#Run Simulation
system.time(result <- parLapply(cl = clust,
                                X = scenarios,
                                fun = sim_multi))

```

```

##      user      system elapsed
##    0.03      0.29 12895.11

```

```

if (!pdp | !ale) {
  print_results(result)
}

```

```

## Setting 1: N = 40 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##           Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1337308 0.02707131 0.6918114
## Mean(s) of simulated LM Variable Effect(s):
##   -1.342806 0.06758416 1.555188
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.8573876 0.8911448 0.9772488 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.867785 0.9686166 1.077175 .
## Number of Smaller Nulls:
##   0 0 0
##
## Setting 2: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##           Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1271024 0.0001112434 1.046716
## Mean(s) of simulated LM Variable Effect(s):
##   -1.460256 -0.002444664 1.650206
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.633569 0.6962919 1.005863 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.650955 0.6988509 1.054374 .
## Number of Smaller Nulls:
##   5 2 0

```

```

##
## Setting 3: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      -0.1188575 -0.02998854 1.074858
## Mean(s) of simulated LM Variable Effect(s):
##      -1.492 -0.003597859 1.645414
## True Variable Effect(s):
##      -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##      0.7018322 0.6079802 0.9902942 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.712211 0.6947659 1.023991 .
## Number of Smaller Nulls:
##      35 31 17
##
## Setting 4: N = 40 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      0.07942526 0.1140769 0.7005192
## Mean(s) of simulated LM Variable Effect(s):
##      -1.275989 0.02057096 1.558429
## True Variable Effect(s):
##      -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##      0.803951 0.8402024 0.9670982 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.8616364 0.8932272 0.9987328 .
## Number of Smaller Nulls:
##      0 0 0
##
## Setting 5: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      -0.008614438 0.0531074 1.015871
## Mean(s) of simulated LM Variable Effect(s):
##      -1.486999 0.001757267 1.621449
## True Variable Effect(s):
##      -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##      0.768818 0.7676243 0.962604 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.8206123 0.8691415 1.04791 .
## Number of Smaller Nulls:
##      1 0 0
##
## Setting 6: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      0.02875332 0.001213341 0.9873439
## Mean(s) of simulated LM Variable Effect(s):
##      -1.501432 -0.0001724951 1.651357
## True Variable Effect(s):
##      -0.02 0 1.00668

```



```

## Standard Error of simulated Variable Effects (RF):
## 0.8016442 0.827408 0.9521504 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.7713477 0.8474571 1.030681 .
## Number of Smaller Nulls:
## 29 22 17
##
## Setting 7: N = 40 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{|\text{abs}(x_2)|} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## -0.4515843 -0.002863822 0.9255397
## Mean(s) of simulated LM Variable Effect(s):
## -1.303185 0.01739952 1.562421
## True Variable Effect(s):
## -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
## 0.3133012 0.3941806 0.3990695 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.3722936 0.385729 0.3969445 .
## Number of Smaller Nulls:
## 0 0 0
##
## Setting 8: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{|\text{abs}(x_2)|} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## -0.4677231 -0.01061008 1.200232
## Mean(s) of simulated LM Variable Effect(s):
## -1.481354 -0.01748415 1.660515
## True Variable Effect(s):
## -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
## 0.2323384 0.2245272 0.2831688 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.2297917 0.2313156 0.2772416 .
## Number of Smaller Nulls:
## 3 0 0
##
## Setting 9: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{|\text{abs}(x_2)|} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## -0.4624744 -0.007644217 1.169605
## Mean(s) of simulated LM Variable Effect(s):
## -1.496184 0.004258761 1.649541
## True Variable Effect(s):
## -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
## 0.192642 0.1915976 0.1958708 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.2026622 0.2105222 0.2225681 .
## Number of Smaller Nulls:
## 19 16 13
##
## Setting 10: N = 40 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{|\text{abs}(x_2)|} \cdot x_2^2 + \exp(x_3)$  ; N_Trees = 2000

```

```

## Mean(s) of simulated RF Variable Effect(s):
##   -0.06779552 0.08652354 0.6180337
## Mean(s) of simulated LM Variable Effect(s):
##   -1.308947 -0.01687951 1.569977
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.3245694 0.3208936 0.3728172 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.3192738 0.3205646 0.3169596 .
## Number of Smaller Nulls:
##   0 0 0
##
## Setting 11: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##   Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.3312557 0.07208273 1.113169
## Mean(s) of simulated LM Variable Effect(s):
##   -1.460666 0.001458446 1.64081
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.2339631 0.2586251 0.2620907 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2556644 0.2589832 0.2676696 .
## Number of Smaller Nulls:
##   0 0 1
##
## Setting 12: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##   Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.4381163 0.006241213 1.164374
## Mean(s) of simulated LM Variable Effect(s):
##   -1.493169 0.009763232 1.641572
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.206711 0.2093721 0.2193499 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2131665 0.2187801 0.2212075 .
## Number of Smaller Nulls:
##   19 12 15

```

```
effect_plots <- plot_effects(result)
```

```

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

```

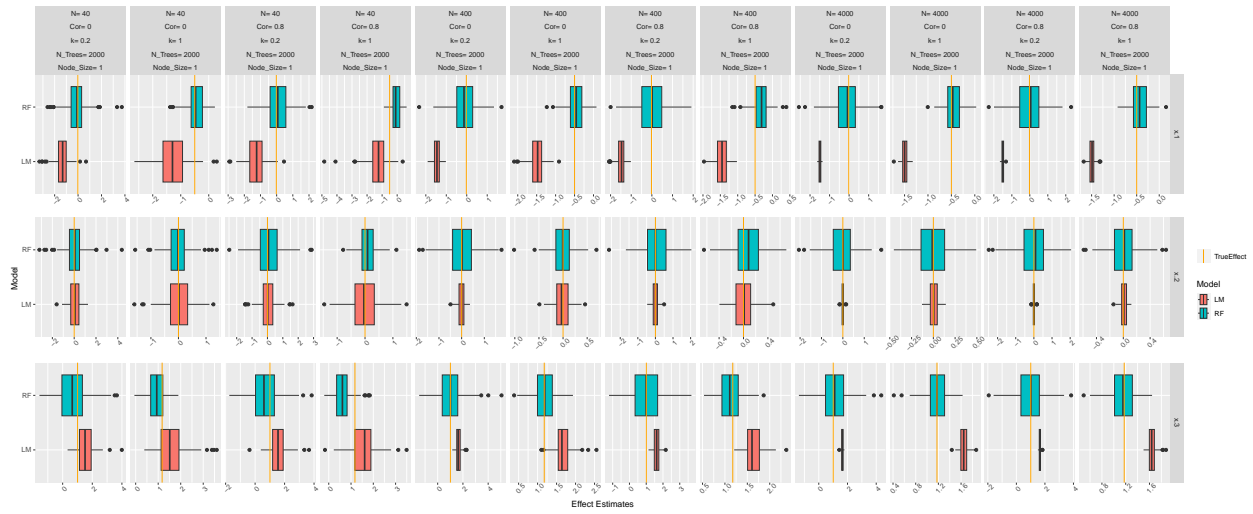
```
se_plot <- plot_se(result)
```

```

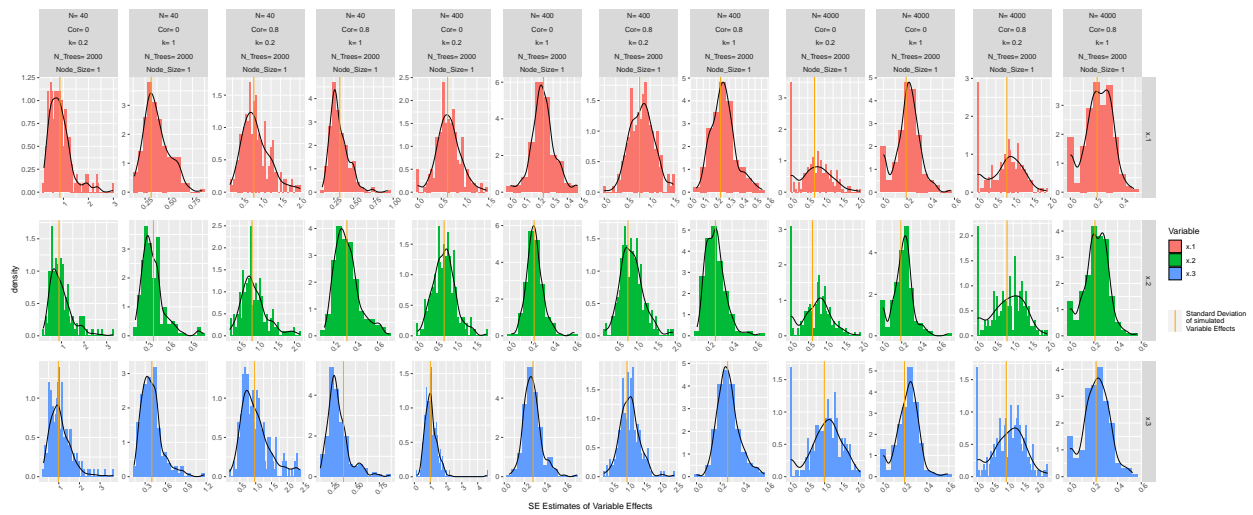
## 'summarise()' has grouped output by 'N', 'cor', 'k', 'num.trees', 'node_size'.
## You can override using the '.groups' argument.

```

## effect\_plots



## se\_plot



```
if (pdp | ale) {
  plot_marginal(result)
}
```

```

n <- c(40, 400, 4000) ; num.trees <- 2000 ; repeats <- 200; cor <- c(0, 0.8)
k <- c(0.2, 1); node_size <- c(1); pdp <- F; ale <- F

formulas <- c("-0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)+2*x.1*x.3")
#parallel::clusterExport(cl = clust, varlist = 'formulas')
scenarios <- data.frame(expand.grid(n, num.trees, formulas, repeats,
                                   cor, k, node_size, pdp, ale))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats",
                       "Correlation", "k", "Node_Size", "pdp", "ale")
scenarios$k_idx <- (scenarios$k == unique(scenarios$k)[1])
scenarios[, "Formula"] <- as.character(scenarios[, "Formula"]) ### Formula became Factor
scenarios <- split(scenarios, seq(nrow(scenarios)))
#Run Simulation
system.time(result <- parLapply(cl = clust,
                               X = scenarios,
                               fun = sim_multi))

```

```

##      user      system elapsed
##    0.08      0.26 18508.89

```

```

if (!pdp | !ale) {
  print_results(result)
}

```

```

## Setting 1: N = 40 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##           Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)+2*x.1*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1944022 -0.06555448 0.6390321 0.6002303
## Mean(s) of simulated LM Variable Effect(s):
##   -1.295749 0.01801668 1.601718 2.054144
## True Variable Effect(s):
##   -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##   0.9458678 1.247689 1.087957 0.8874048 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.9784848 1.237399 1.082395 5.468003 .
## Number of Smaller Nulls:
##   0 0 0 0
##
## Setting 2: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##           Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)+2*x.1*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.2236597 -0.01099025 0.8371436 1.61722
## Mean(s) of simulated LM Variable Effect(s):
##   -1.459294 -0.01361833 1.632673 2.027881
## True Variable Effect(s):
##   -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##   0.6650026 0.7178183 0.8400925 1.589784 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.7287025 0.6872502 0.9567501 3.300323 .
## Number of Smaller Nulls:
##   5 8 1 0

```

```

##
## Setting 3: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \times x_1^3 + 0.5 \times \sqrt{\text{abs}(x_2)} \times x_2^2 + \exp(x_3) + 2 \times x_1 \times x_3$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      -0.157432 0.02881343 0.9744015 2.162223
## Mean(s) of simulated LM Variable Effect(s):
##      -1.500532 -0.005805026 1.640649 1.985818
## True Variable Effect(s):
##      -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##      0.8175803 0.6838058 0.852806 2.918796 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.7792034 0.6963114 0.9248402 3.775035 .
## Number of Smaller Nulls:
##      31 33 18 8
##
## Setting 4: N = 40 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \times x_1^3 + 0.5 \times \sqrt{\text{abs}(x_2)} \times x_2^2 + \exp(x_3) + 2 \times x_1 \times x_3$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      -0.07116219 0.09191754 0.4892991 0.3133947
## Mean(s) of simulated LM Variable Effect(s):
##      -1.311128 -0.0451286 1.475052 2.150955
## True Variable Effect(s):
##      -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##      0.801983 0.8955242 0.8509487 0.7667671 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.8863797 0.961317 0.8860581 4.365884 .
## Number of Smaller Nulls:
##      0 0 0 0
##
## Setting 5: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \times x_1^3 + 0.5 \times \sqrt{\text{abs}(x_2)} \times x_2^2 + \exp(x_3) + 2 \times x_1 \times x_3$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      -0.1781402 0.05923261 0.9310203 1.23444
## Mean(s) of simulated LM Variable Effect(s):
##      -1.463312 -0.004643284 1.640657 2.138719
## True Variable Effect(s):
##      -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##      0.8418857 0.8429652 0.9534604 1.866773 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.8767944 0.8838802 0.9958003 3.617957 .
## Number of Smaller Nulls:
##      2 0 0 0
##
## Setting 6: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##      Formula =  $-0.5 \times x_1^3 + 0.5 \times \sqrt{\text{abs}(x_2)} \times x_2^2 + \exp(x_3) + 2 \times x_1 \times x_3$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##      -0.114025 0.03147855 0.9978046 1.814822
## Mean(s) of simulated LM Variable Effect(s):
##      -1.492539 -0.003186332 1.64289 2.158889
## True Variable Effect(s):
##      -0.02 0 1.00668 2

```

```

## Standard Error of simulated Variable Effects (RF):
## 0.8052797 0.7871338 0.912355 2.652645 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.8790537 0.7920526 1.021791 3.874219 .
## Number of Smaller Nulls:
## 26 35 14 7
##
## Setting 7: N = 40 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3) + 2 \cdot x_1 \cdot x_3$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## -0.4931883 -0.02080687 0.8665076 0.4125638
## Mean(s) of simulated LM Variable Effect(s):
## -1.32159 -0.04164891 1.625615 2.004655
## True Variable Effect(s):
## -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
## 0.4377962 0.4894399 0.4957618 0.1642021 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.4647725 0.4804675 0.4631037 0.3816909 .
## Number of Smaller Nulls:
## 0 0 0 0
##
## Setting 8: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3) + 2 \cdot x_1 \cdot x_3$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## -0.5546311 -0.005233235 1.140729 1.388085
## Mean(s) of simulated LM Variable Effect(s):
## -1.485769 0.01343367 1.647931 1.96719
## True Variable Effect(s):
## -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
## 0.28036 0.2424009 0.2644414 0.1859668 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.2693828 0.2406729 0.291062 0.2397367 .
## Number of Smaller Nulls:
## 1 4 1 0
##
## Setting 9: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3) + 2 \cdot x_1 \cdot x_3$  ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
## -0.5222473 0.003652105 1.206589 1.869659
## Mean(s) of simulated LM Variable Effect(s):
## -1.500434 0.005467198 1.645676 1.996379
## True Variable Effect(s):
## -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
## 0.2095719 0.1812864 0.2374324 0.1678025 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.2326599 0.187102 0.2299038 0.2137602 .
## Number of Smaller Nulls:
## 17 29 17 7
##
## Setting 10: N = 40 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
## Formula =  $-0.5 \cdot x_1^3 + 0.5 \cdot \sqrt{\text{abs}(x_2)} \cdot x_2^2 + \exp(x_3) + 2 \cdot x_1 \cdot x_3$  ; N_Trees = 2000

```

```

## Mean(s) of simulated RF Variable Effect(s):
##   -0.1070715 0.1490711 0.5755318 0.2486971
## Mean(s) of simulated LM Variable Effect(s):
##   -1.245464 0.03637977 1.465996 2.042665
## True Variable Effect(s):
##   -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
##   0.5630911 0.4666433 0.5161534 0.1116244 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.5074894 0.4435319 0.456842 0.2955154 .
## Number of Smaller Nulls:
##   0 0 0 0
##
## Setting 11: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##   Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)+2*x.1*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.4344778 0.08920062 1.049625 0.8561358
## Mean(s) of simulated LM Variable Effect(s):
##   -1.472562 -0.007833496 1.608846 2.16004
## True Variable Effect(s):
##   -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
##   0.3112466 0.2894948 0.3263943 0.1567921 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.3037381 0.2817964 0.3046525 0.2110884 .
## Number of Smaller Nulls:
##   0 1 0 0
##
## Setting 12: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##   Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)+2*x.1*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   -0.5225611 -0.001819847 1.149229 1.433151
## Mean(s) of simulated LM Variable Effect(s):
##   -1.497736 -0.003462416 1.643604 2.15028
## True Variable Effect(s):
##   -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
##   0.222855 0.2301844 0.2328526 0.1525023 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2329096 0.2242668 0.2254405 0.188377 .
## Number of Smaller Nulls:
##   19 16 22 13

```

```
effect_plots <- plot_effects(result)
```

```

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

```

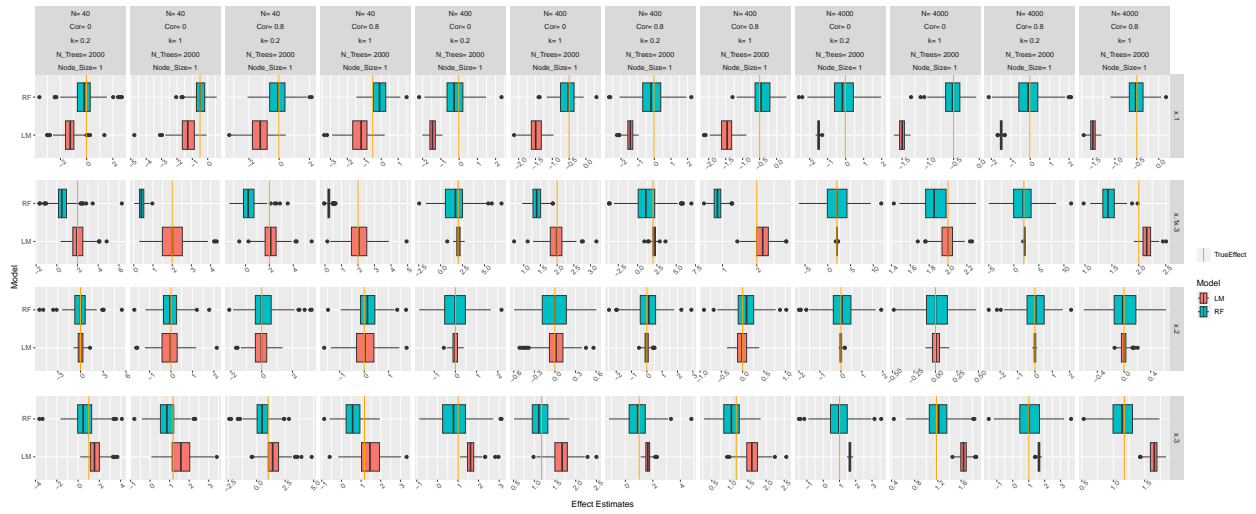
```
se_plot <- plot_se(result)
```

```

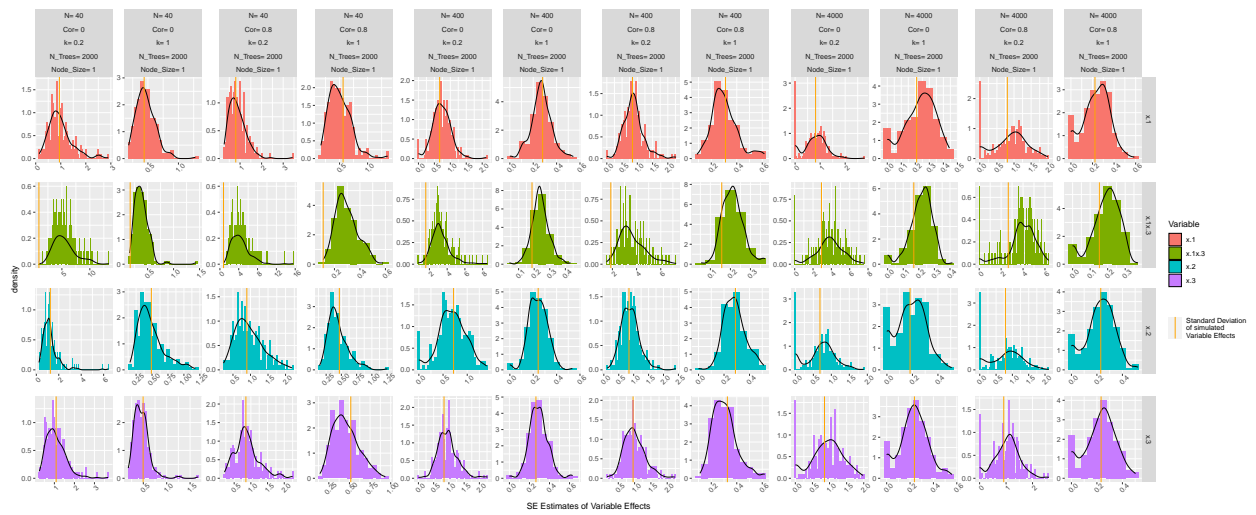
## 'summarise()' has grouped output by 'N', 'cor', 'k', 'num.trees', 'node_size'.
## You can override using the '.groups' argument.

```

## effect\_plots



## se\_plot



```
if (pdp | ale) {
  plot_marginal(result)
}
```