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

Felix Kapulla

```
knitr::opts_chunk$set(fig.width=14, fig.height=8)
```

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

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

# cores <- detectCores()
# clust <- makeCluster(cores-1)
# parallel::clusterEvalQ(clust,
#                        expr = {source('C:/Users/feix_/iCloudDrive/Studium Master/CQM - Thesis Interns
```

Simulation

```
n <- c(400, 4000) ; num.trees <- 2000 ; repeats <- 250; cor <- c(0, 0.8); k <- c(0.2, 1)
formulas <- c("2*x.1+4*x.2-0.5*x.3+2.2*x.4")
scenarios <- data.frame(expand.grid(n, num.trees, formulas, repeats, cor, k))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats", "Correlation", "k")
scenarios[,"Formula"] <- as.character(scenarios[,"Formula"]) ### Formula became Factor
scenarios <- split(scenarios, seq(nrow(scenarios)))

system.time(result <- lapply(X = scenarios, FUN = sim_multi))
```

```
##     user   system  elapsed
## 36305.78  1686.36 11183.97
```

```
#Run Simulation
# system.time(result <- parLapply(cl = clust,
#                                 X = scenarios,
#                                 fun = sim_multi))

#stopCluster(clust)
```

```
print_results(result)
```

```
## Setting: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.965209 4.267526 -0.3320597 2.09078
## Mean(s) of simulated LM Variable Effect(s):
##   1.997715 4.004991 -0.5036113 2.197786
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.9431981 1.423855 0.3042736 1.057268 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.9818118 1.58754 0.3098638 1.088154 .
## Number of Smaller Nulls:
##   6 1 64 5
##
## Setting: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   2.040327 4.266912 -0.4040833 2.343346
## Mean(s) of simulated LM Variable Effect(s):
##   2.000391 3.998798 -0.5014079 2.198219
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.7515774 1.02902 0.3321001 0.9234399 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.9422311 1.069438 0.4458704 0.9923229 .
## Number of Smaller Nulls:
##   38 39 64 36
##
## Setting: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.81864 3.981484 -0.04534022 2.053035
## Mean(s) of simulated LM Variable Effect(s):
##   2.002126 3.997283 -0.4917747 2.193697
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.9460376 1.207295 0.3502132 0.9018823 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.009462 1.293247 0.3937666 1.039801 .
## Number of Smaller Nulls:
##   2 0 20 0
##
## Setting: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.949942 3.865862 -0.3282933 2.19587
## Mean(s) of simulated LM Variable Effect(s):
##   2.000537 4.001059 -0.4979449 2.196312
```

```
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.8760926 0.9962621 0.3960802 0.8602232 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.9315534 1.003257 0.4890873 0.9290456 .
## Number of Smaller Nulls:
##   35 25 55 35
##
## Setting: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.766932 3.871763 -0.2926801 2.00035
## Mean(s) of simulated LM Variable Effect(s):
##   1.998499 3.9987 -0.4950683 2.197246
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.2827285 0.2990587 0.1278144 0.2831958 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.3022563 0.3827534 0.1500172 0.3217181 .
## Number of Smaller Nulls:
##   7 1 31 5
##
## Setting: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.944626 3.986083 -0.378251 2.142014
## Mean(s) of simulated LM Variable Effect(s):
##   2.000194 3.999311 -0.5013327 2.19948
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.219015 0.2390971 0.1423793 0.2159049 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2412912 0.2535133 0.1677845 0.2223871 .
## Number of Smaller Nulls:
##   26 26 48 42
##
## Setting: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.690992 3.712089 0.03443495 1.871452
## Mean(s) of simulated LM Variable Effect(s):
##   2.006064 4.005409 -0.509882 2.199054
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.2594629 0.3076533 0.1504277 0.2364912 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2877417 0.3477457 0.1700364 0.2953543 .
## Number of Smaller Nulls:
##   7 4 9 4
##
```

```
## Setting: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.845824 3.846657 -0.225316 2.056398
## Mean(s) of simulated LM Variable Effect(s):
##   1.999023 3.997147 -0.5000075 2.202507
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.2246142 0.2024879 0.1494592 0.2265333 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2330888 0.2363462 0.153077 0.2213793 .
## Number of Smaller Nulls:
##   37 29 61 33
```

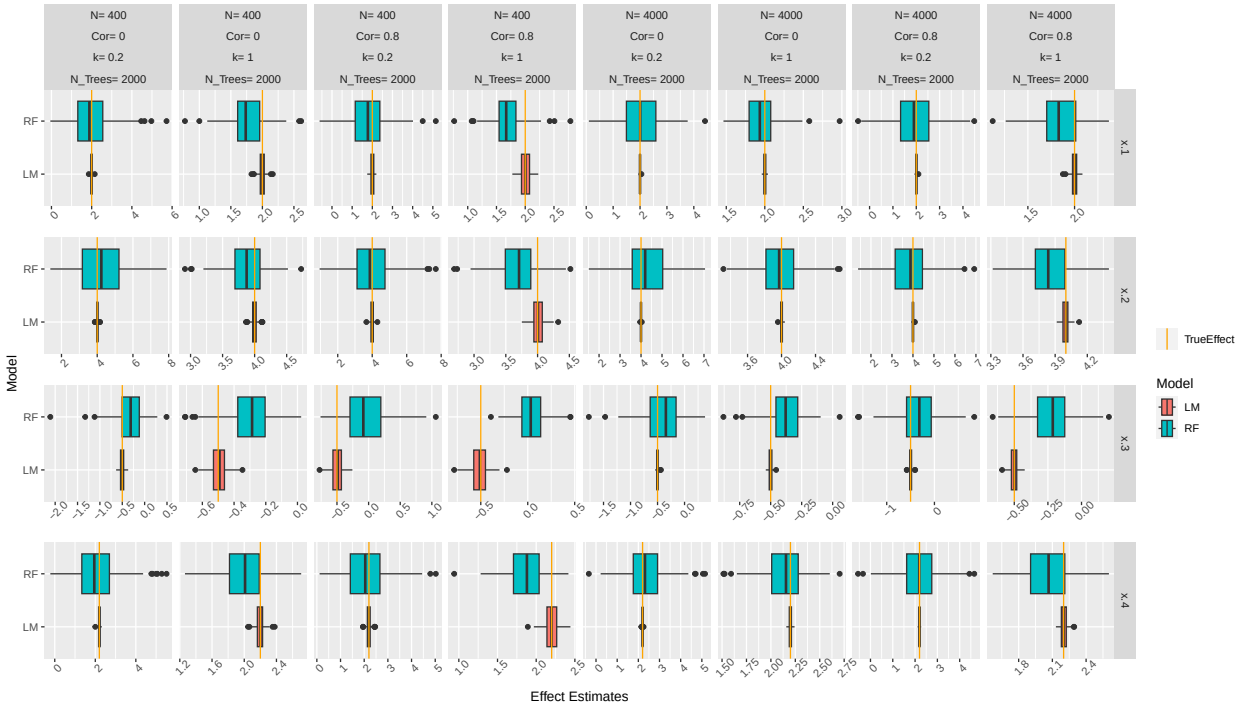
```
effect_plots <- plot_effects(result)
```

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

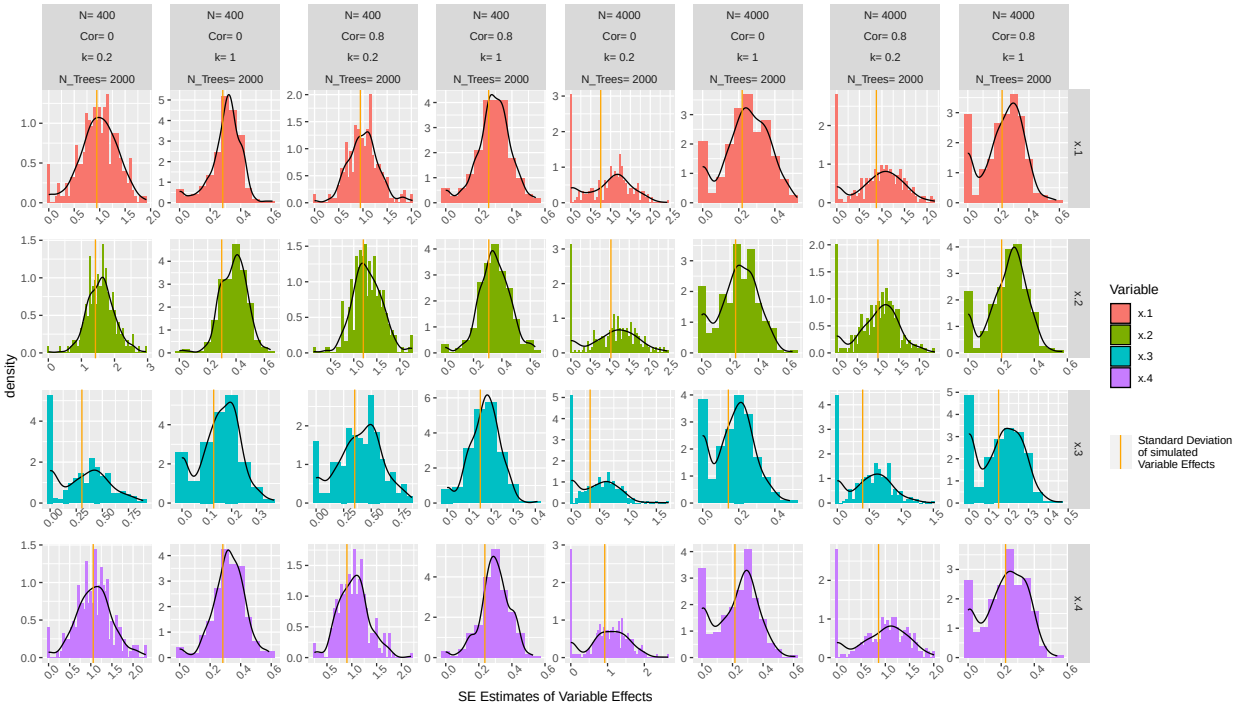
```
se_plot <- plot_se(result)
```

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

```
effect_plots
```



```
se_plot
```


N= 400 Cor= 0 k= 0.2 N_Trees= 2000
N= 400 Cor= 0 k= 1 N_Trees= 2000
N= 400 Cor= 0.8 k= 0.2 N_Trees= 2000
N= 400 Cor= 0.8 k= 1 N_Trees= 2000
N= 4000 Cor= 0 k= 0.2 N_Trees= 2000
N= 4000 Cor= 0 k= 1 N_Trees= 2000
N= 4000 Cor= 0.8 k= 0.2 N_Trees= 2000
N= 4000 Cor= 0.8 k= 1 N_Trees= 2000
x.1
x.2
x.3
x.4
density
SE Estimates of Variable Effects
Variable
x.1
x.2
x.3
x.4
Standard Deviation of simulated Variable Effects

```r
###### Simulation Setup
n <- c(400, 4000) ; num.trees <- 2000 ; repeats <- 250; cor <- c(0, 0.8); k <- c(0.2, 1)
formulas <- "-0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)"
scenarios <- data.frame(expand.grid(n, num.trees, formulas, repeats, cor, k))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats", "Correlation", "k")
scenarios[,"Formula"] <- as.character(scenarios[,"Formula"]) ### Formula became Factor
scenarios <- split(scenarios, seq(nrow(scenarios)))

system.time(result <- lapply(X = scenarios, FUN = sim_multi))
```

```
##     user   system  elapsed
## 23815.66   919.06  6699.50
```

```r
print_results(result)
```

```
## Setting: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.05517508 -0.03375279 0.9932003
## Mean(s) of simulated LM Variable Effect(s):
##   -1.473776 -0.006540286 1.642859
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.570317 0.568601 0.9133347 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.5690727 0.6052748 0.9568411 .
## Number of Smaller Nulls:
##   5 6 0
##
## Setting: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.04151893 0.01350635 0.9942633
## Mean(s) of simulated LM Variable Effect(s):
##   -1.496828 -0.0002781891 1.651095
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.601245 0.5977239 0.8887756 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6797113 0.6063174 0.9699674 .
## Number of Smaller Nulls:
##   34 39 20
##
## Setting: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ;
##          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):
##   -0.01821721 -0.04705758 0.9106827
## Mean(s) of simulated LM Variable Effect(s):
##   -1.462113 -0.001481522 1.63265
## True Variable Effect(s):
```

```
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.7047826 0.7586626 0.9602875 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.7549346 0.7684505 0.9561389 .
## Number of Smaller Nulls:
##   0 3 0
##
## Setting: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ;
##          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):
##   -0.01312315 -0.02808391 1.075928
## Mean(s) of simulated LM Variable Effect(s):
##   -1.498381 0.002266096 1.651848
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.693425 0.6845035 0.9133091 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.7953873 0.7333242 0.9429067 .
## Number of Smaller Nulls:
##   20 26 15
##
## Setting: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.4372079 -0.0035957 1.196761
## Mean(s) of simulated LM Variable Effect(s):
##   -1.477606 -0.01679932 1.642683
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.1938584 0.2320574 0.2540093 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2172871 0.2272902 0.2512716 .
## Number of Smaller Nulls:
##   1 0 2
##
## Setting: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.4879365 0.01991488 1.193119
## Mean(s) of simulated LM Variable Effect(s):
##   -1.49556 0.002571554 1.643043
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.2040286 0.1949091 0.1837892 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2030853 0.1964358 0.196296 .
## Number of Smaller Nulls:
##   23 17 20
##
## Setting: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ;
```

```
##          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):
##   -0.3646537 0.07616817 1.03456
## Mean(s) of simulated LM Variable Effect(s):
##   -1.465562 0.002018818 1.629855
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.218371 0.2485473 0.245178 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2322144 0.2437497 0.246937 .
## Number of Smaller Nulls:
##   0 0 0
##
## Setting: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ;
##          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):
##   -0.4570394 0.004881383 1.147194
## Mean(s) of simulated LM Variable Effect(s):
##   -1.501591 0.001360155 1.648133
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.205943 0.2052303 0.2111963 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2057468 0.2015092 0.206319 .
## Number of Smaller Nulls:
##   18 17 20
```

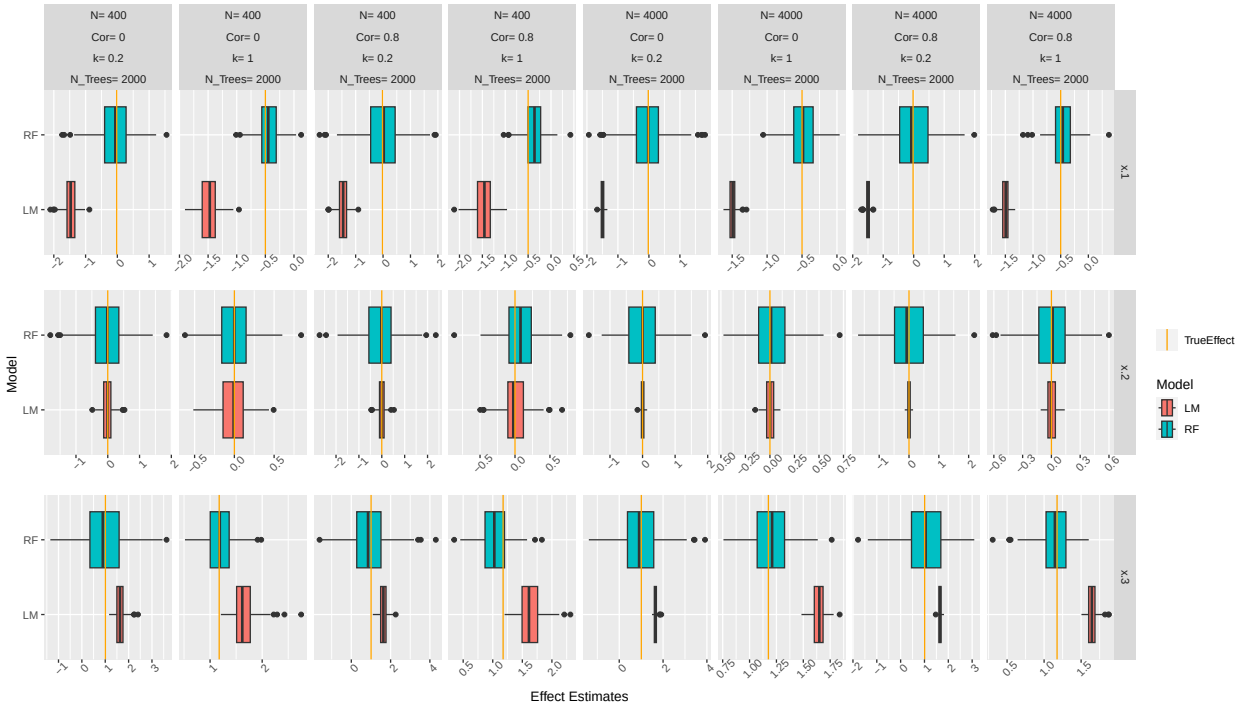
```
effect_plots <- plot_effects(result)
```

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

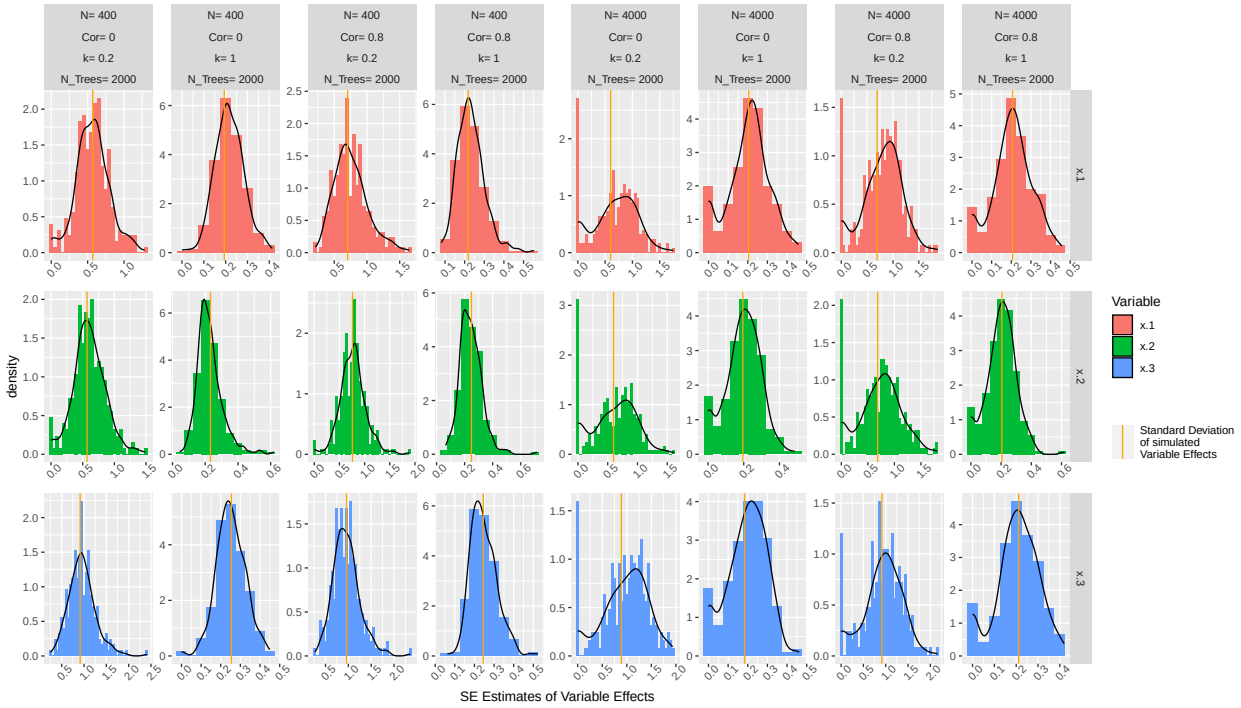
```
se_plot <- plot_se(result)
```

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

```
effect_plots
```


N= 400
Cor= 0
k= 0.2
N_Trees= 2000
N= 400
Cor= 0
k= 1
N_Trees= 2000
N= 400
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 400
Cor= 0.8
k= 1
N_Trees= 2000
N= 4000
Cor= 0
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0
k= 1
N_Trees= 2000
N= 4000
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0.8
k= 1
N_Trees= 2000
RF
LM
x.1
x.2
x.3
Model
TrueEffect
Model
LM
RF
Effect Estimates

```
se_plot
```


N= 400
Cor= 0
k= 0.2
N_Trees= 2000
N= 400
Cor= 0
k= 1
N_Trees= 2000
N= 400
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 400
Cor= 0.8
k= 1
N_Trees= 2000
N= 4000
Cor= 0
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0
k= 1
N_Trees= 2000
N= 4000
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0.8
k= 1
N_Trees= 2000
density
x.1
x.2
x.3
Variable
x.1
x.2
x.3
Standard Deviation
of simulated
Variable Effects
SE Estimates of Variable Effects

```r
###### Simulation Setup
n <- c(400, 4000) ; num.trees <- 2000 ; repeats <- 250; cor <- c(0, 0.8); k <- c(0.2, 1)
formulas <- "-0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3)+2*x.1*x.3"
scenarios <- data.frame(expand.grid(n, num.trees, formulas, repeats, cor, k))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats", "Correlation", "k")
scenarios[,"Formula"] <- as.character(scenarios[,"Formula"]) ### Formula became Factor
scenarios <- split(scenarios, seq(nrow(scenarios)))

system.time(result <- lapply(X = scenarios, FUN = sim_multi))
```

```
##     user   system  elapsed
## 24326.14  1045.21  7623.89
```

```r
print_results(result)
```

```
## Setting: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.2617215 -0.01303212 0.8729003 -1.24544
## Mean(s) of simulated LM Variable Effect(s):
##   -1.476232 0.006820613 1.628672 -1.977952
## True Variable Effect(s):
##   -0.02 0 1.00668 -2
## Standard Error of simulated Variable Effects (RF):
##   0.7110667 0.5752855 0.8483684 1.521983 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6763248 0.6416681 0.8742279 3.116641 .
## Number of Smaller Nulls:
##   8 10 4 0
##
## Setting: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.1051088 0.07283378 1.01868 -1.702606
## Mean(s) of simulated LM Variable Effect(s):
##   -1.495528 0.007392282 1.646669 -1.999474
## True Variable Effect(s):
##   -0.02 0 1.00668 -2
## Standard Error of simulated Variable Effects (RF):
##   0.6910596 0.5213594 0.8574914 2.535203 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6962795 0.6672613 0.9077305 3.438026 .
## Number of Smaller Nulls:
##   42 31 23 7
##
## Setting: N = 400 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ;
##          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):
##   -0.118696 -0.02061344 0.9206186 -1.334671
## Mean(s) of simulated LM Variable Effect(s):
##   -1.483769 0.02040566 1.620507 -2.104549
## True Variable Effect(s):
```

```
##   -0.02 0 1.00668 -2
## Standard Error of simulated Variable Effects (RF):
##   0.7269616 0.8184177 0.8598017 1.597498 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.7581673 0.8262042 0.9111209 3.16857 .
## Number of Smaller Nulls:
##   3 0 0 0
##
## Setting: N = 4000 ; k = 0.2 N_Trees = 2000 ; Correlation = 0.8 ;
##          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):
##   -0.02539144 0.09613184 0.9182025 -1.689418
## Mean(s) of simulated LM Variable Effect(s):
##   -1.498882 -0.004729353 1.652181 -2.141808
## True Variable Effect(s):
##   -0.02 0 1.00668 -2
## Standard Error of simulated Variable Effects (RF):
##   0.7746123 0.7001957 0.8919061 2.438265 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.7746518 0.7368573 0.940207 3.544391 .
## Number of Smaller Nulls:
##   33 30 18 6
##
## Setting: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.5767496 -0.008863908 1.147194 -1.367873
## Mean(s) of simulated LM Variable Effect(s):
##   -1.472233 0.003002541 1.663962 -1.999171
## True Variable Effect(s):
##   -0.5 0 1.175201 -2
## Standard Error of simulated Variable Effects (RF):
##   0.272281 0.2460968 0.2778249 0.181773 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2647262 0.2516951 0.2786153 0.2357523 .
## Number of Smaller Nulls:
##   1 2 0 0
##
## Setting: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0 ;
##          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):
##   -0.5140143 0.01420247 1.191345 -1.874147
## Mean(s) of simulated LM Variable Effect(s):
##   -1.497948 -0.0007648589 1.650086 -1.999316
## True Variable Effect(s):
##   -0.5 0 1.175201 -2
## Standard Error of simulated Variable Effects (RF):
##   0.2022275 0.1641536 0.2312354 0.1433263 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.1964457 0.1729708 0.2199082 0.189914 .
## Number of Smaller Nulls:
##   37 36 24 15
##
## Setting: N = 400 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ;
```

```
##          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):
##   -0.4786726 0.1022427 1.112126 -0.8224673
## Mean(s) of simulated LM Variable Effect(s):
##   -1.462129 -0.007270924 1.631927 -2.123536
## True Variable Effect(s):
##   -0.5 0 1.175201 -2
## Standard Error of simulated Variable Effects (RF):
##   0.2909293 0.255398 0.2988391 0.1460866 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2900593 0.2582096 0.2880865 0.1960932 .
## Number of Smaller Nulls:
##   0 0 1 0
##
## Setting: N = 4000 ; k = 1 N_Trees = 2000 ; Correlation = 0.8 ;
##          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):
##   -0.5138036 0.06104696 1.156546 -1.418713
## Mean(s) of simulated LM Variable Effect(s):
##   -1.495227 0.003528658 1.64752 -2.138025
## True Variable Effect(s):
##   -0.5 0 1.175201 -2
## Standard Error of simulated Variable Effects (RF):
##   0.233506 0.2124944 0.2268767 0.1437127 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2293417 0.2036662 0.2194627 0.1842217 .
## Number of Smaller Nulls:
##   13 26 18 11
```

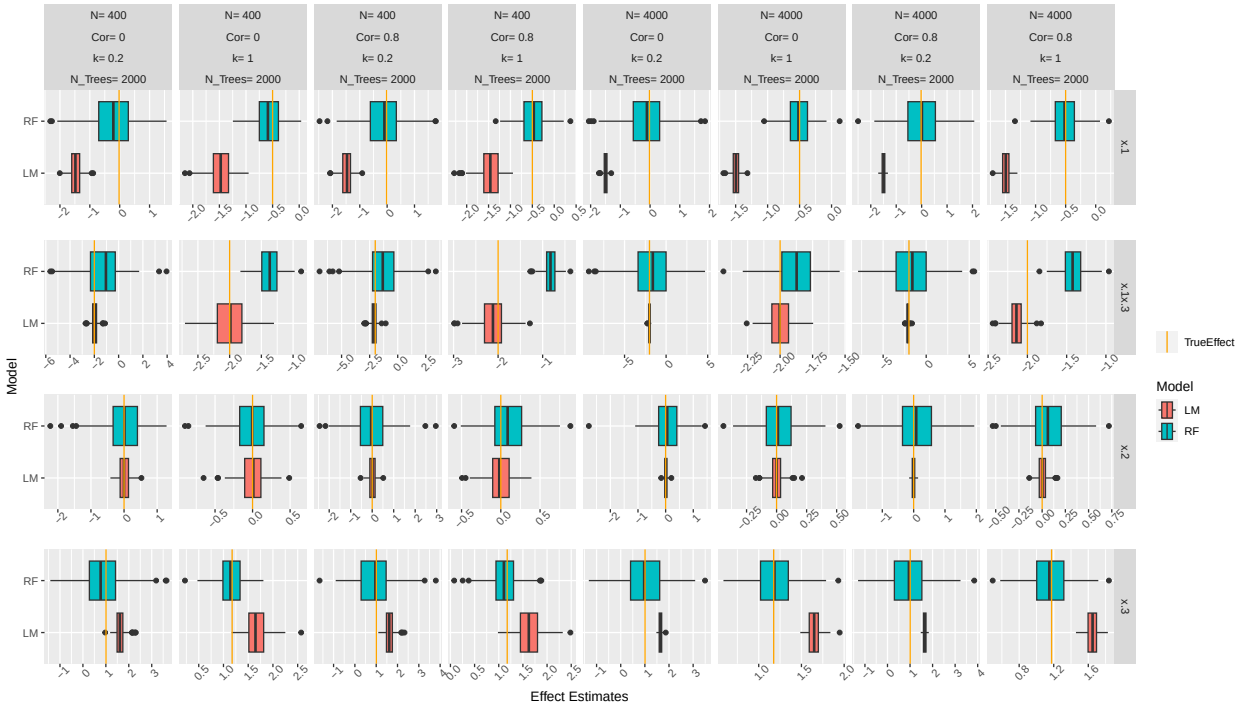
```
effect_plots <- plot_effects(result)
```

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

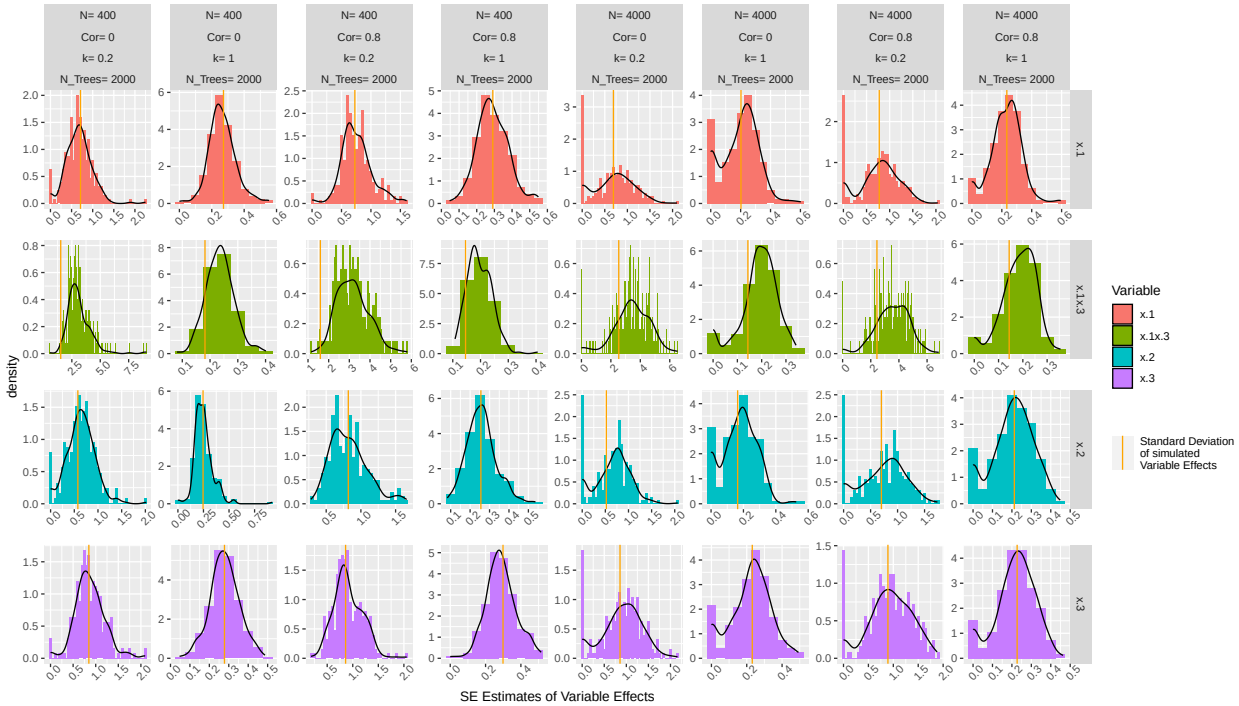
```
se_plot <- plot_se(result)
```

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

```
effect_plots
```


N= 400
Cor= 0
k= 0.2
N_Trees= 2000
N= 400
Cor= 0
k= 1
N_Trees= 2000
N= 400
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 400
Cor= 0.8
k= 1
N_Trees= 2000
N= 4000
Cor= 0
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0
k= 1
N_Trees= 2000
N= 4000
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0.8
k= 1
N_Trees= 2000
RF
LM
x.1
x.1x.3
x.2
x.3
Model
Effect Estimates
TrueEffect
Model
LM
RF

```
se_plot
```


N= 400
Cor= 0
k= 0.2
N_Trees= 2000
N= 400
Cor= 0
k= 1
N_Trees= 2000
N= 400
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 400
Cor= 0.8
k= 1
N_Trees= 2000
N= 4000
Cor= 0
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0
k= 1
N_Trees= 2000
N= 4000
Cor= 0.8
k= 0.2
N_Trees= 2000
N= 4000
Cor= 0.8
k= 1
N_Trees= 2000
x.1
x.1x.3
x.2
x.3
density
SE Estimates of Variable Effects
Variable
x.1
x.1x.3
x.2
x.3
Standard Deviation
of simulated
Variable Effects