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(40, 50) ; num.trees <- 20 ; repeats <- 25; cor <- c(0, 0.8)
k <- c(0.2, 1); node_size <- 5
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, node_size))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats",
                        "Correlation", "k", "Node_Size")
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
##   43.03    9.84   39.96
```

```
#Run Simulation
# system.time(result <- parLapply(cl = clust,
```

```
#                                   X = scenarios,
#                                   fun = sim_multi))

#stopCluster(clust)
```

```
print_results(result)
```

```
## Setting 1: N = 40 ; k = 0.2 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.293039 2.823353 -0.1146227 0.9463268
## Mean(s) of simulated LM Variable Effect(s):
##   1.994159 3.975377 -0.4879226 2.185778
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   1.579979 2.344078 0.7911009 1.313768 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   3.67758 5.0524 3.144025 3.075179 .
## Number of Smaller Nulls:
##   4 3 5 4
##
## Setting 2: N = 50 ; k = 0.2 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.959121 3.917349 -0.08754782 0.9916058
## Mean(s) of simulated LM Variable Effect(s):
##   1.990526 3.982539 -0.48767 2.147575
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   1.67414 3.163388 0.8974997 1.311645 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   5.111942 6.028743 3.370979 4.121695 .
## Number of Smaller Nulls:
##   2 3 2 4
##
## Setting 3: N = 40 ; k = 0.2 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   2.082655 3.451465 0.3071516 1.662583
## Mean(s) of simulated LM Variable Effect(s):
##   2.058299 3.81628 -0.3868705 2.234876
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   1.869002 3.260487 1.102627 1.177605 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   3.840427 2.689241 2.808322 3.047182 .
## Number of Smaller Nulls:
##   3 7 3 5
##
## Setting 4: N = 50 ; k = 0.2 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
```

```
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.845686 2.891804 0.4329536 1.548244
## Mean(s) of simulated LM Variable Effect(s):
##   1.978147 4.07235 -0.4860833 2.212241
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   1.486736 1.763718 0.5445254 1.213892 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   3.970326 4.862646 2.493792 3.402092 .
## Number of Smaller Nulls:
##   2 1 2 4
##
## Setting 5: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.242048 3.264768 -0.1138687 1.177231
## Mean(s) of simulated LM Variable Effect(s):
##   2.020705 3.966643 -0.494279 2.16944
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.5758784 0.6940417 0.4716733 0.5124453 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.458856 2.209826 1.104625 1.400523 .
## Number of Smaller Nulls:
##   2 1 4 6
##
## Setting 6: N = 50 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.407063 3.190766 -0.2136661 1.279258
## Mean(s) of simulated LM Variable Effect(s):
##   2.033857 3.973116 -0.509954 2.181575
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.5377443 0.6051727 0.4952127 0.4518675 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.633568 1.729065 1.27542 1.331243 .
## Number of Smaller Nulls:
##   2 3 4 2
##
## Setting 7: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.781818 2.842144 0.6824515 1.543753
## Mean(s) of simulated LM Variable Effect(s):
##   2.068748 3.945048 -0.6017114 2.249828
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.5413725 0.5045085 0.5191765 0.6215922 .
```

```
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.475732 1.861409 1.136683 1.298215 .
## Number of Smaller Nulls:
##   5 1 5 4
##
## Setting 8: N = 50 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.750779 3.282151 0.433323 1.90105
## Mean(s) of simulated LM Variable Effect(s):
##   1.999726 4.024675 -0.6007898 2.291425
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.6653212 0.6947989 0.4166605 0.7977166 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.610353 1.632623 0.8088956 1.260199 .
## Number of Smaller Nulls:
##   3 4 3 5
```

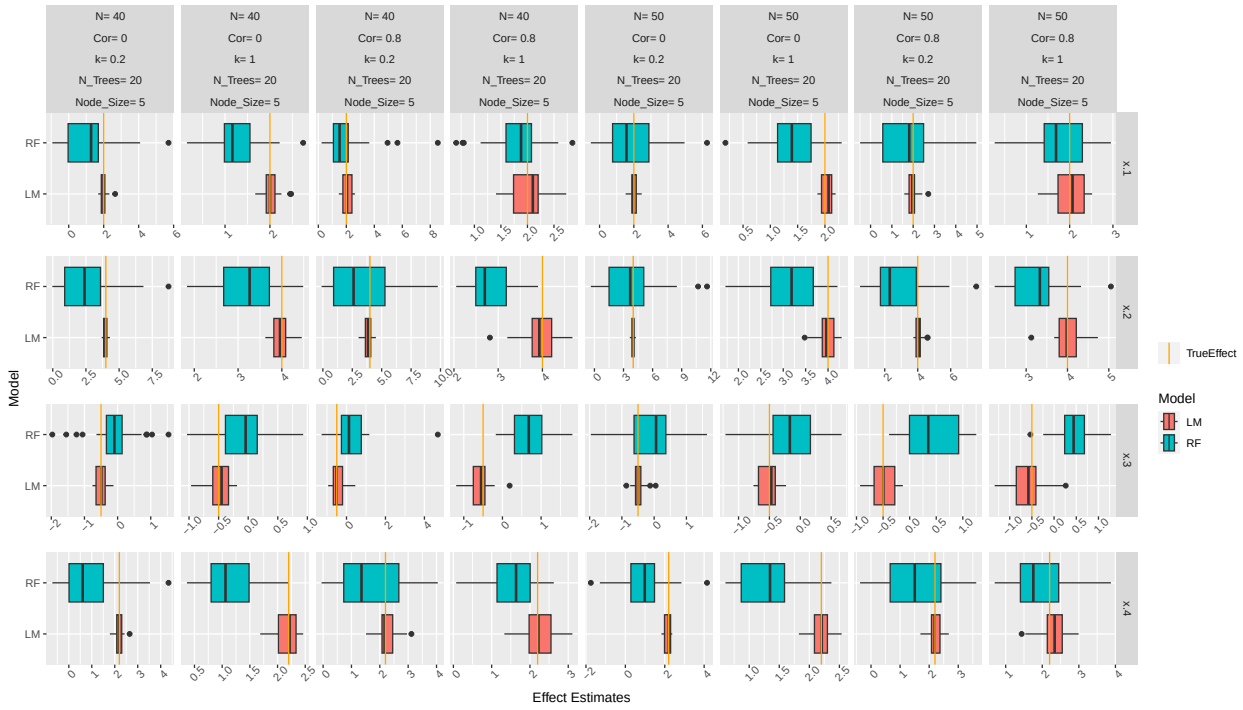
```
effect_plots <- plot_effects(result)
```

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

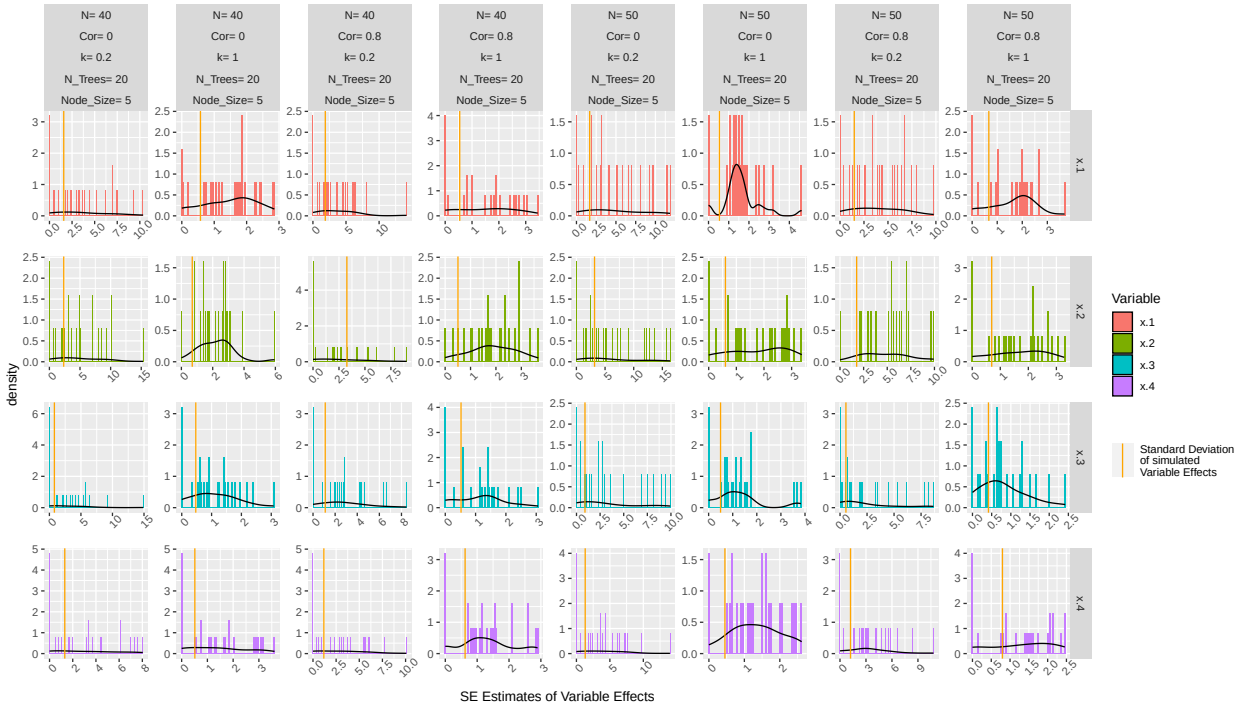
```
se_plot <- plot_se(result)
```

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

```
effect_plots
```

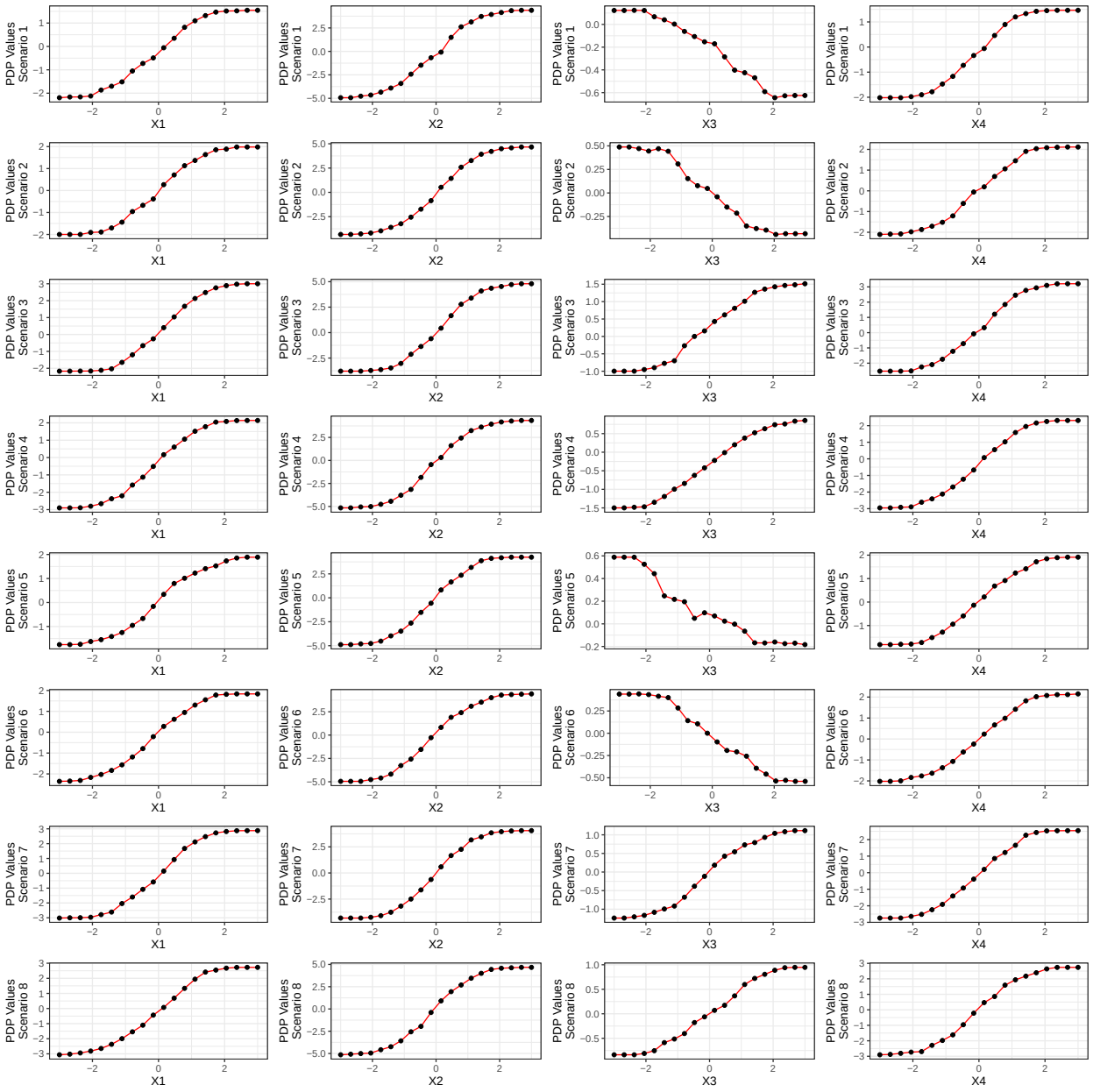


```
se_plot
```




```
plot_pdps(result)
```

```
## Warning: ‘aes_string()‘ was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation ideoms with ‘aes()‘
```


PDP Values Scenario 1
PDP Values Scenario 2
PDP Values Scenario 3
PDP Values Scenario 4
PDP Values Scenario 5
PDP Values Scenario 6
PDP Values Scenario 7
PDP Values Scenario 8
X1
X2
X3
X4

```
n <- c(40) ; num.trees <- 20 ; repeats <- 25; cor <- c(0, 0.8)
k <- c(1); node_size <- c(1, 5, 100)
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, node_size))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats",
                        "Correlation", "k", "Node_Size")
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
##   29.25    7.99   27.83
```

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

#stopCluster(clust)
```

```
print_results(result)
```

```
## Setting 1: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 1 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.104038 3.495612 -0.1587173 1.599868
## Mean(s) of simulated LM Variable Effect(s):
##   1.930083 3.967827 -0.5304181 2.260975
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.5688011 0.5405794 0.6173563 0.8042276 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.610025 1.913942 1.426369 1.80761 .
## Number of Smaller Nulls:
##   1 2 3 2
##
## Setting 2: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 1 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.747985 3.019059 0.6200091 1.541359
## Mean(s) of simulated LM Variable Effect(s):
##   2.040781 4.187374 -0.598852 2.072962
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.7500497 0.702475 0.4818204 0.7018552 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.710343 1.626629 0.8273573 1.368364 .
## Number of Smaller Nulls:
```

```
##   2 6 7 4
##
## Setting 3: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.163356 3.277097 -0.2383031 1.268407
## Mean(s) of simulated LM Variable Effect(s):
##   1.987141 4.035647 -0.4570431 2.229726
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.6879358 0.8328212 0.3229346 0.6087237 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.50899 1.464756 1.285189 1.470751 .
## Number of Smaller Nulls:
##   5 4 3 2
##
## Setting 4: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   1.948803 3.014864 0.5928555 1.728028
## Mean(s) of simulated LM Variable Effect(s):
##   2.101058 4.107949 -0.6875868 2.138372
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0.5200333 0.7634527 0.4612539 0.5744274 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.655989 1.559578 1.076759 1.578225 .
## Number of Smaller Nulls:
##   3 8 3 3
##
## Setting 5: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 100 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   0 0 0 0
## Mean(s) of simulated LM Variable Effect(s):
##   2.006184 4.06102 -0.4507606 2.150255
## True Variable Effect(s):
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0 0 0 0 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0 0 0 0 .
## Number of Smaller Nulls:
##   0 0 0 0
##
## Setting 6: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 100 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3+2.2*x.4 ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   0 0 0 0
## Mean(s) of simulated LM Variable Effect(s):
##   2.144336 4.031344 -0.4030524 1.993509
## True Variable Effect(s):
```

```
##   2 4 -0.5 2.2
## Standard Error of simulated Variable Effects (RF):
##   0 0 0 0 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0 0 0 0 .
## Number of Smaller Nulls:
##   0 0 0 0
```

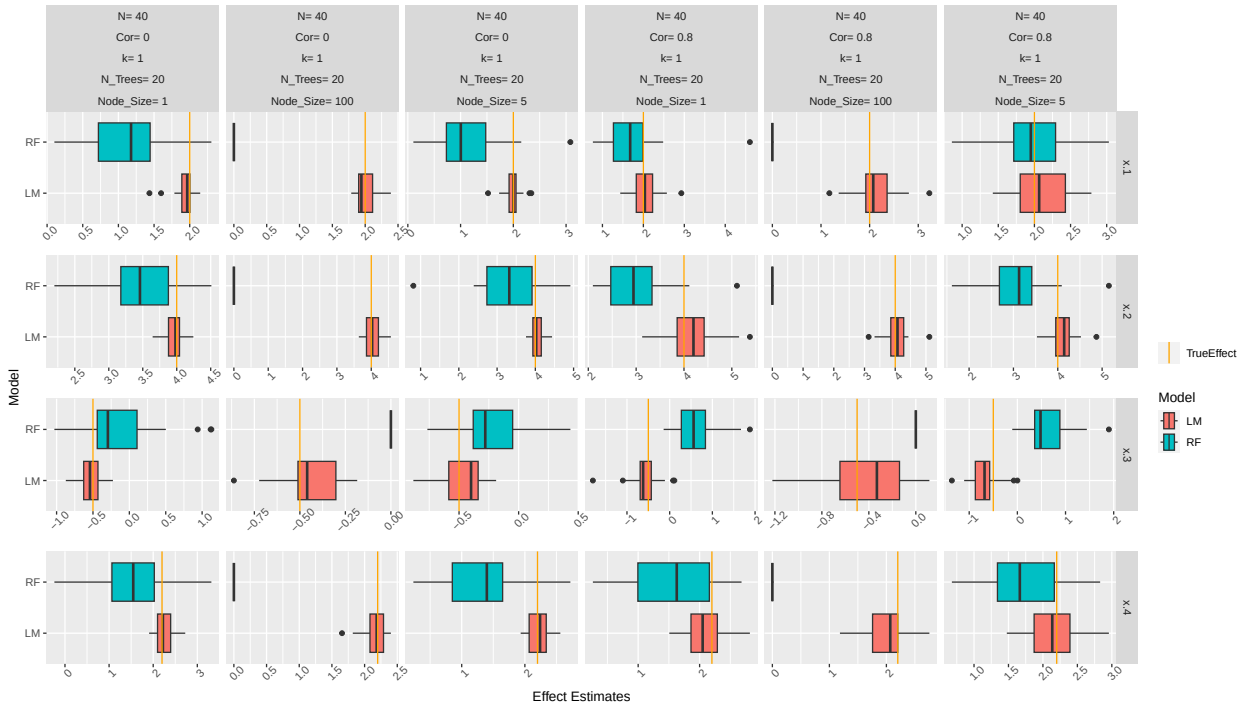
```
effect_plots <- plot_effects(result)
```

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

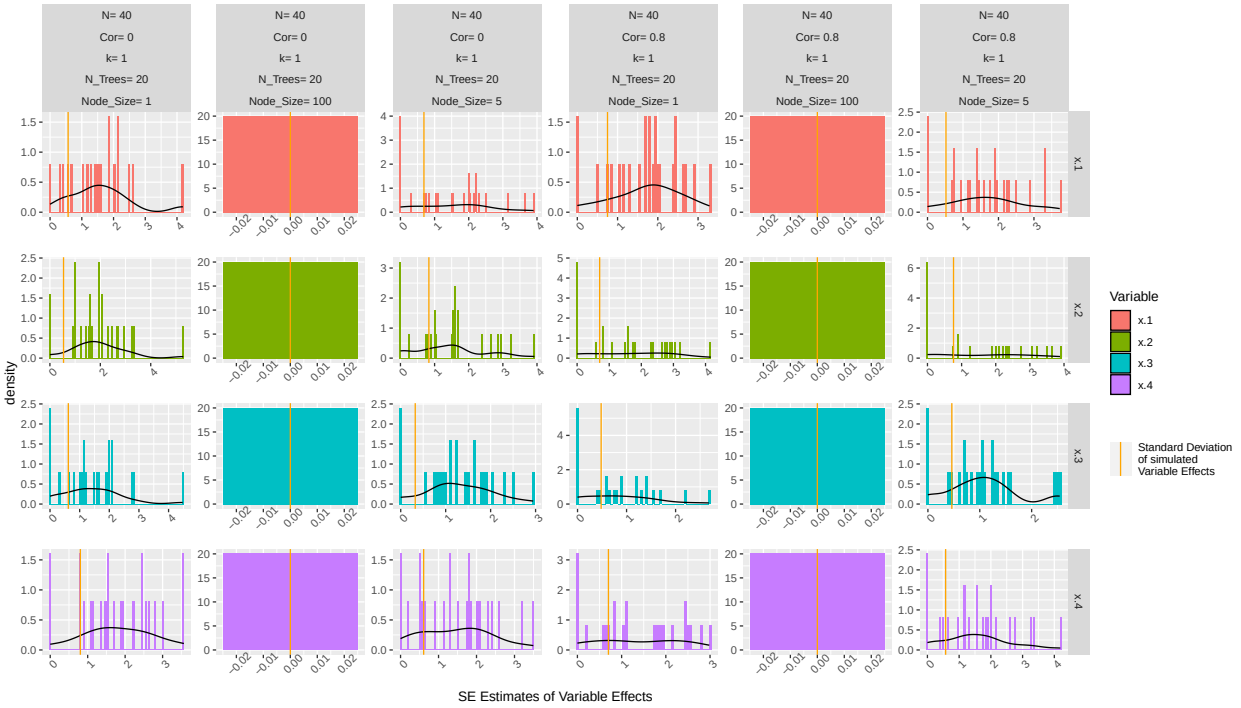
```
se_plot <- plot_se(result)
```

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

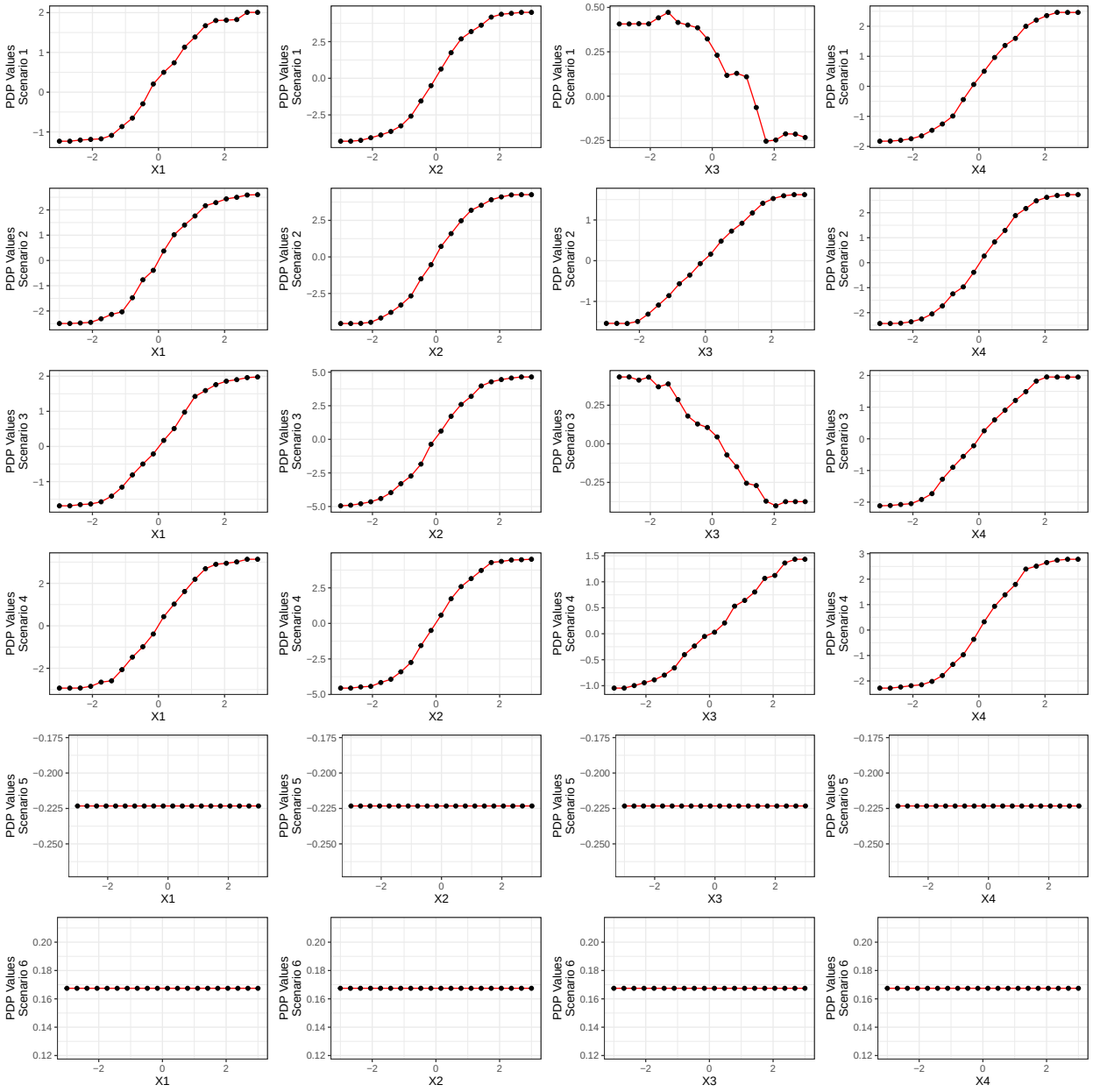
```
effect_plots
```




```
se_plot
```




```
plot_pdps(result)
```


PDP Values Scenario 1
PDP Values Scenario 2
PDP Values Scenario 3
PDP Values Scenario 4
PDP Values Scenario 5
PDP Values Scenario 6
X1
X2
X3
X4

```
###### Simulation Setup
n <- c(40, 50) ; num.trees <- 20 ; repeats <- 25; cor <- c(0, 0.8)
k <- c(0.2, 1); node_size <- 5
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, node_size))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats",
                        "Correlation", "k", "Node_Size")
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
##   24.67    6.24   23.36
```

```
print_results(result)
```

```
## Setting 1: N = 40 ; k = 0.2 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.2383065 0.5035213 0.8017567
## Mean(s) of simulated LM Variable Effect(s):
##   -1.215276 0.0138097 1.609443
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   1.158414 1.638047 1.421525 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   2.931499 4.110165 2.895838 .
## Number of Smaller Nulls:
##   1 3 1
##
## Setting 2: N = 50 ; k = 0.2 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   0.05076738 0.4598203 0.782606
## Mean(s) of simulated LM Variable Effect(s):
##   -1.371802 0.119551 1.559593
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   1.096733 0.8584174 1.229146 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   2.248172 2.683015 2.685024 .
## Number of Smaller Nulls:
##   4 2 3
##
## Setting 3: N = 40 ; k = 0.2 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   0.08358869 0.4244149 0.6840983
```

```
## Mean(s) of simulated LM Variable Effect(s):
##   -1.085987 -0.01113633 1.447611
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.9027144 0.7990724 0.956671 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.720131 1.840557 2.322291 .
## Number of Smaller Nulls:
##   2 3 3
##
## Setting 4: N = 50 ; k = 0.2 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1822505 0.1547169 0.472977
## Mean(s) of simulated LM Variable Effect(s):
##   -1.457731 -0.02049567 1.629576
## True Variable Effect(s):
##   -0.02 0 1.00668
## Standard Error of simulated Variable Effects (RF):
##   0.8299171 0.8635106 0.9567575 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   2.09472 1.632363 1.760703 .
## Number of Smaller Nulls:
##   0 4 3
##
## Setting 5: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.3914309 -0.01263535 0.9766338
## Mean(s) of simulated LM Variable Effect(s):
##   -1.290413 -0.0772748 1.585941
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.3116721 0.3404323 0.4730807 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.9595078 0.5656035 0.8551153 .
## Number of Smaller Nulls:
##   0 4 4
##
## Setting 6: N = 50 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.5275358 0.02871807 1.034379
## Mean(s) of simulated LM Variable Effect(s):
##   -1.420968 0.09601928 1.725523
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.3880857 0.3570313 0.3871678 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.9137738 0.5890157 0.8122788 .
## Number of Smaller Nulls:
```

```
##   4 8 3
##
## Setting 7: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1286018 0.06445113 0.6419147
## Mean(s) of simulated LM Variable Effect(s):
##   -1.219744 0.1366904 1.481739
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.3376952 0.4058104 0.3644666 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6980748 0.678487 0.7465983 .
## Number of Smaller Nulls:
##   0 1 2
##
## Setting 8: N = 50 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          Formula = -0.5*x.1^3+0.5*sqrt(abs(x.2))*x.2^2+exp(x.3) ; N_Trees = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1616516 0.09274589 0.6952217
## Mean(s) of simulated LM Variable Effect(s):
##   -1.411717 -0.04018101 1.714942
## True Variable Effect(s):
##   -0.5 0 1.175201
## Standard Error of simulated Variable Effects (RF):
##   0.2988537 0.3538586 0.3941551 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.7481734 0.7022945 0.6610528 .
## Number of Smaller Nulls:
##   1 3 3
```

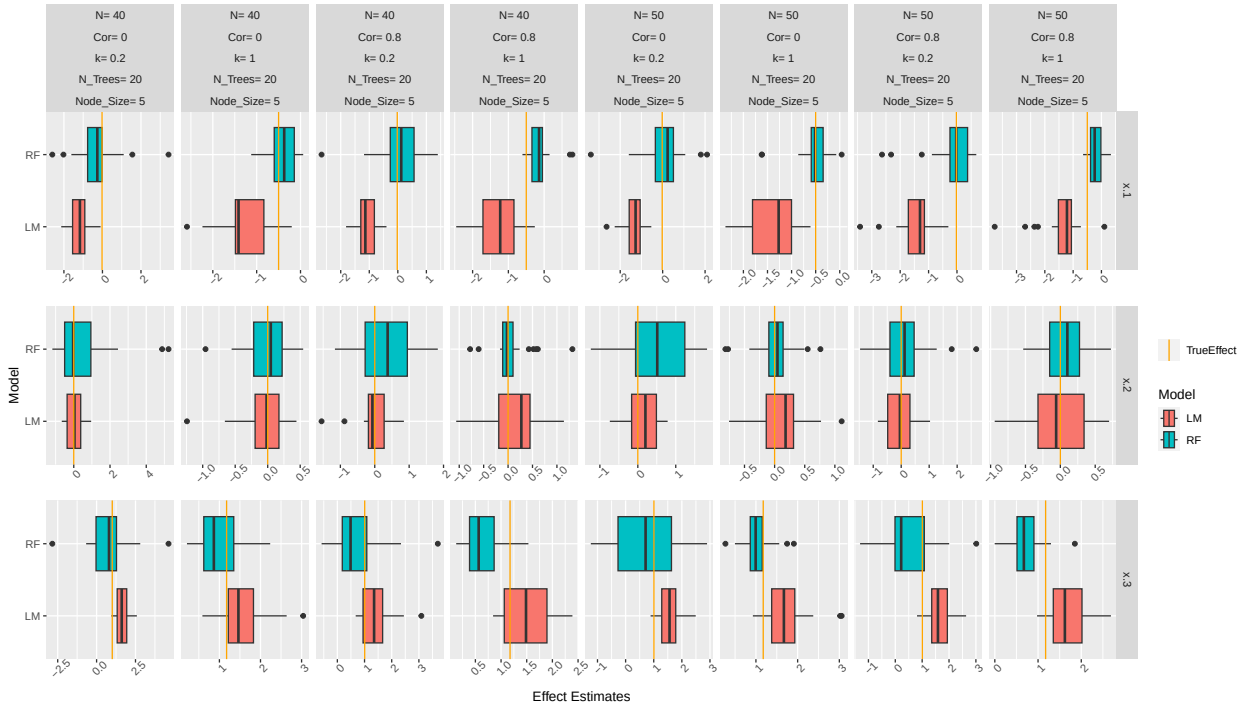
```
effect_plots <- plot_effects(result)
```

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

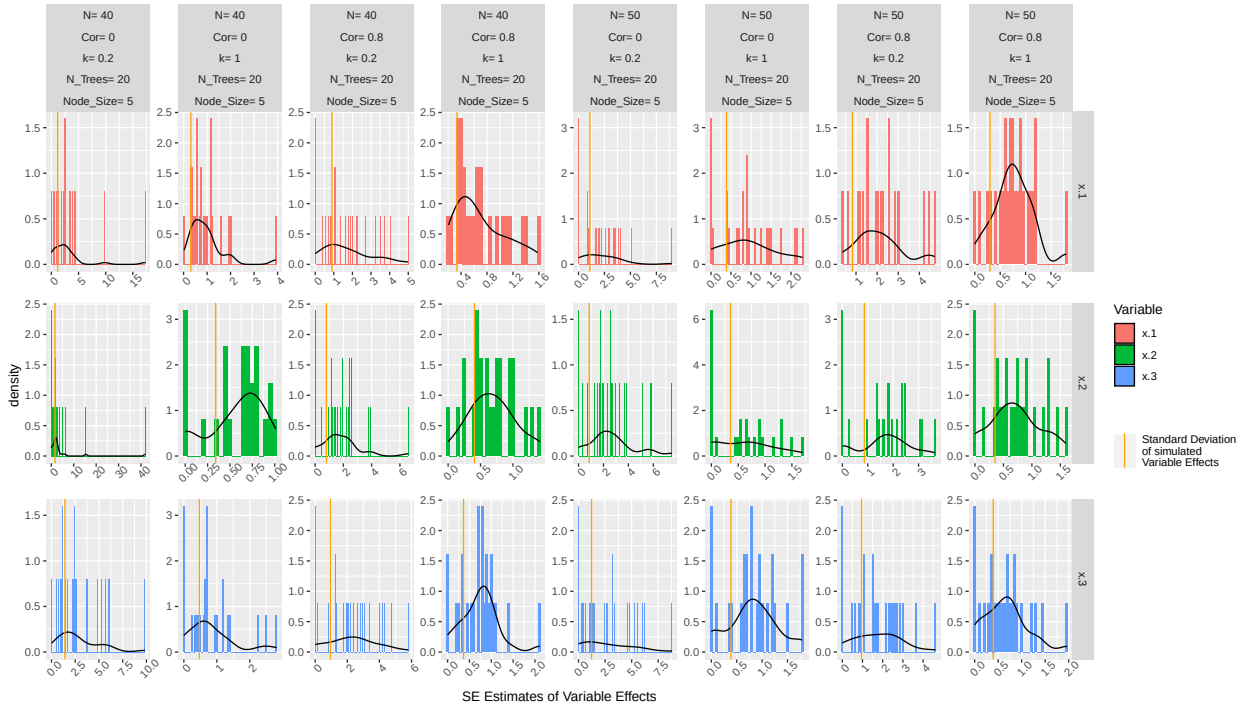
```
se_plot <- plot_se(result)
```

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

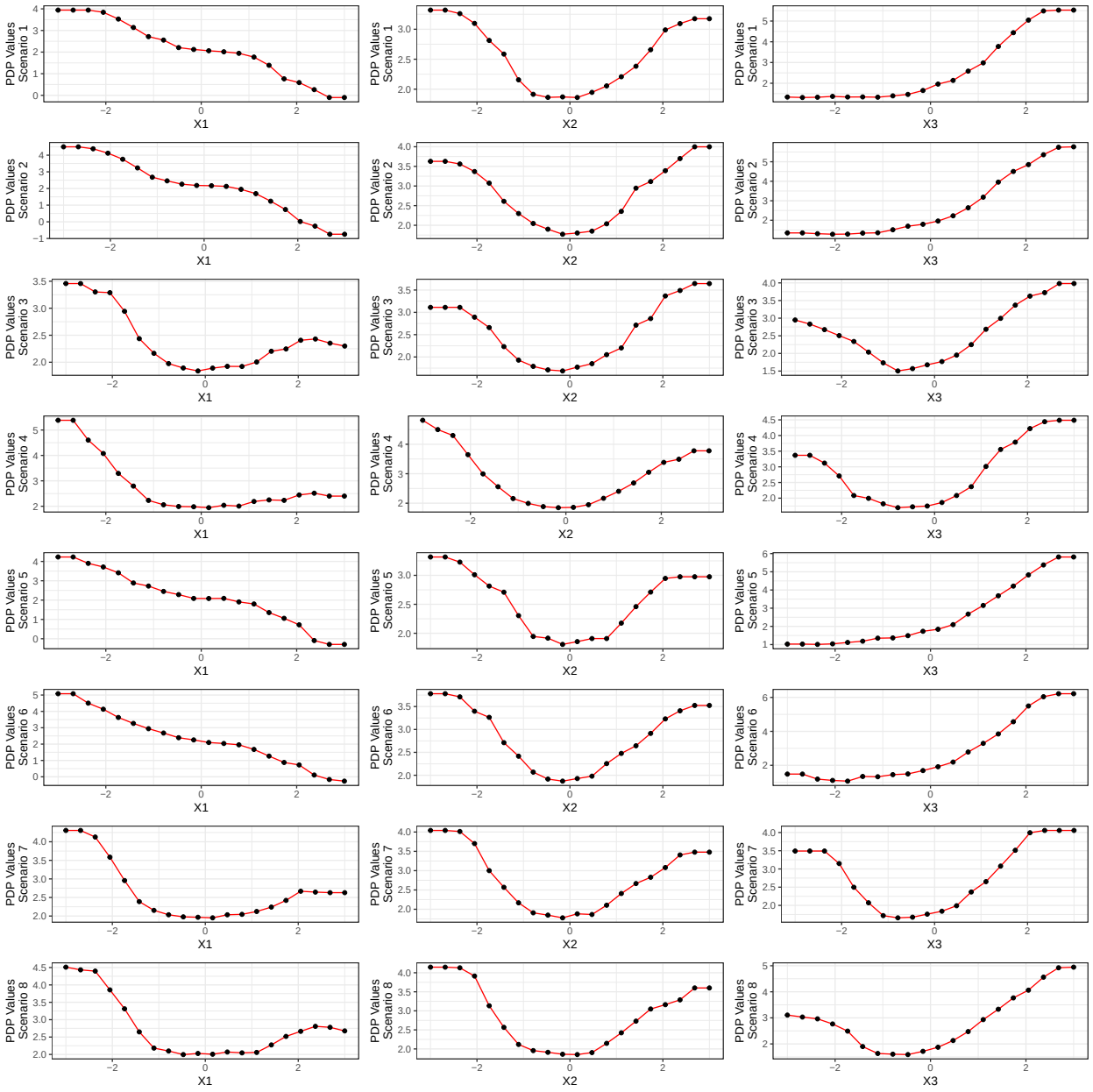
```
effect_plots
```


N= 40
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
RF
LM
x.1
x.2
x.3
Model
Effect Estimates
TrueEffect
Model
LM
RF

```
se_plot
```


N= 40
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
x.1
x.2
x.3
density
SE Estimates of Variable Effects
Variable
x.1
x.2
x.3
Standard Deviation
of simulated
Variable Effects

```
plot_pdps(result)
```



```r
###### Simulation Setup
n <- c(40, 50) ; num.trees <- 20 ; repeats <- 25; cor <- c(0, 0.8)
k <- c(0.2, 1); node_size <- 5
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, node_size))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats",
                        "Correlation", "k", "Node_Size")
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
##   32.09    8.97   30.73
```

```r
print_results(result)
```

```
## Setting 1: N = 40 ; k = 0.2 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.06110768 0.1425864 0.7393097 0.1581627
## Mean(s) of simulated LM Variable Effect(s):
##   -1.297473 0.1119184 1.822333 1.949063
## True Variable Effect(s):
##   -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##   0.9102698 1.339593 1.310429 1.532284 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   2.100648 3.066571 2.79378 8.814775 .
## Number of Smaller Nulls:
##   3 3 1 0
##
## Setting 2: N = 50 ; k = 0.2 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1681967 -0.04315605 0.5025634 0.5021005
## Mean(s) of simulated LM Variable Effect(s):
##   -1.159124 0.06424749 1.462835 1.736191
## True Variable Effect(s):
##   -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##   1.011005 1.219203 0.8939824 1.19136 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   2.454313 3.627714 2.391772 14.87853 .
## Number of Smaller Nulls:
##   2 4 5 0
##
## Setting 3: N = 40 ; k = 0.2 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   0.1733018 0.245918 0.5140827 0.2716799
```

```
## Mean(s) of simulated LM Variable Effect(s):
##   -1.115052 0.1259597 1.787258 2.166452
## True Variable Effect(s):
##   -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##   1.31195 0.7978806 0.9597843 0.753992 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.49018 1.836226 1.992424 9.094957 .
## Number of Smaller Nulls:
##   5 1 3 0
##
## Setting 4: N = 50 ; k = 0.2 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.08534745 -0.3120031 0.4781037 0.2780439
## Mean(s) of simulated LM Variable Effect(s):
##   -1.362247 -0.01990762 1.558649 2.248542
## True Variable Effect(s):
##   -0.02 0 1.00668 2
## Standard Error of simulated Variable Effects (RF):
##   0.7971358 0.8751288 0.7526938 0.8021984 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.989428 2.126182 2.060361 7.693937 .
## Number of Smaller Nulls:
##   1 4 2 0
##
## Setting 5: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.6181199 -0.06566926 0.807537 0.3635091
## Mean(s) of simulated LM Variable Effect(s):
##   -1.683524 -0.03405734 1.649782 1.941387
## True Variable Effect(s):
##   -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
##   0.633718 0.4798979 0.5925669 0.220928 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.8598815 0.8569875 1.134247 0.8665579 .
## Number of Smaller Nulls:
##   5 4 3 0
##
## Setting 6: N = 50 ; k = 1 N_Trees = 20 ; Correlation = 0 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.6963369 0.05845286 0.7981749 0.4233458
## Mean(s) of simulated LM Variable Effect(s):
##   -1.481958 -0.03714207 1.549865 2.035557
## True Variable Effect(s):
##   -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
##   0.702174 0.4305467 0.4794553 0.2679349 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.319975 1.030619 0.9320465 1.074167 .
## Number of Smaller Nulls:
```

```
##   3 2 2 0
##
## Setting 7: N = 40 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1072577 0.1559176 0.4876706 0.2117701
## Mean(s) of simulated LM Variable Effect(s):
##   -1.232711 0.01413605 1.281938 2.249299
## True Variable Effect(s):
##   -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
##   0.547329 0.6325728 0.5868044 0.1000524 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.8789712 0.8904056 0.9019305 0.7472006 .
## Number of Smaller Nulls:
##   4 3 4 0
##
## Setting 8: N = 50 ; k = 1 N_Trees = 20 ; Correlation = 0.8 ; Minimum Node Size = 5 ;
##          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 = 20
## Mean(s) of simulated RF Variable Effect(s):
##   -0.1866816 0.1078551 0.6781268 0.2224004
## Mean(s) of simulated LM Variable Effect(s):
##   -1.588676 0.02199794 1.689844 2.041412
## True Variable Effect(s):
##   -0.5 0 1.175201 2
## Standard Error of simulated Variable Effects (RF):
##   0.5147973 0.4417016 0.5008556 0.1604356 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   1.28058 0.9416196 1.111341 0.8094376 .
## Number of Smaller Nulls:
##   2 4 2 0
```

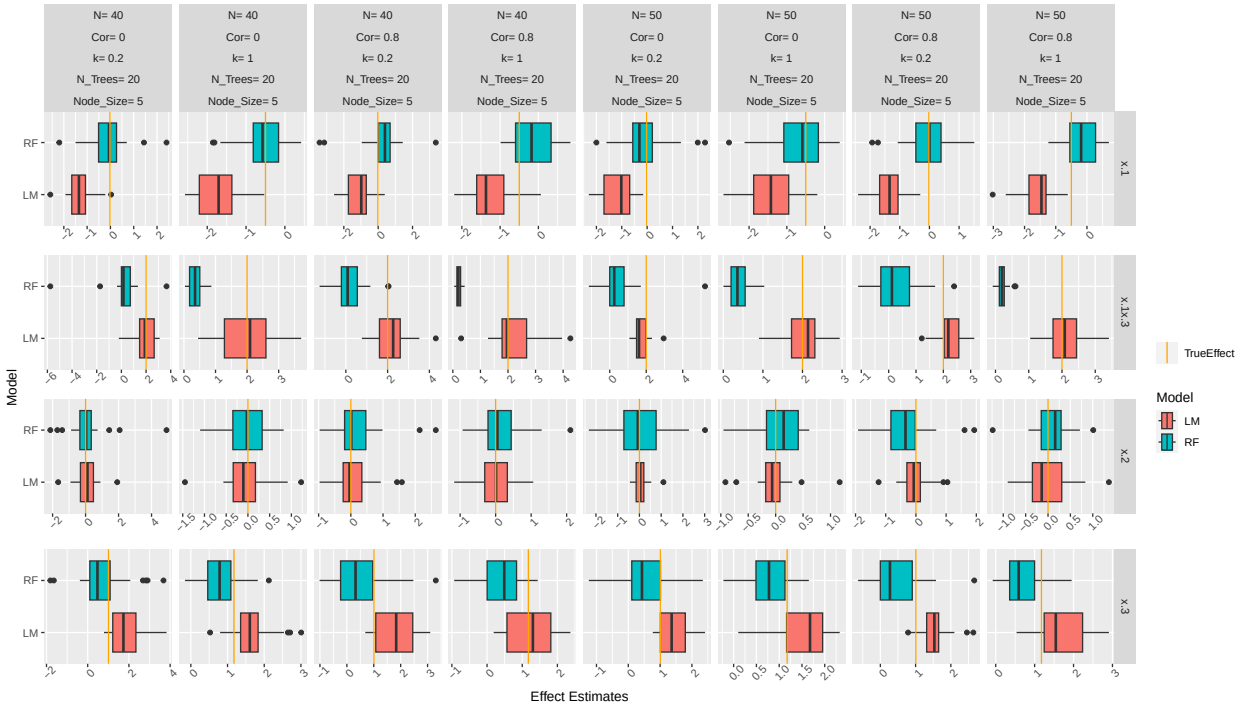
```
effect_plots <- plot_effects(result)
```

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

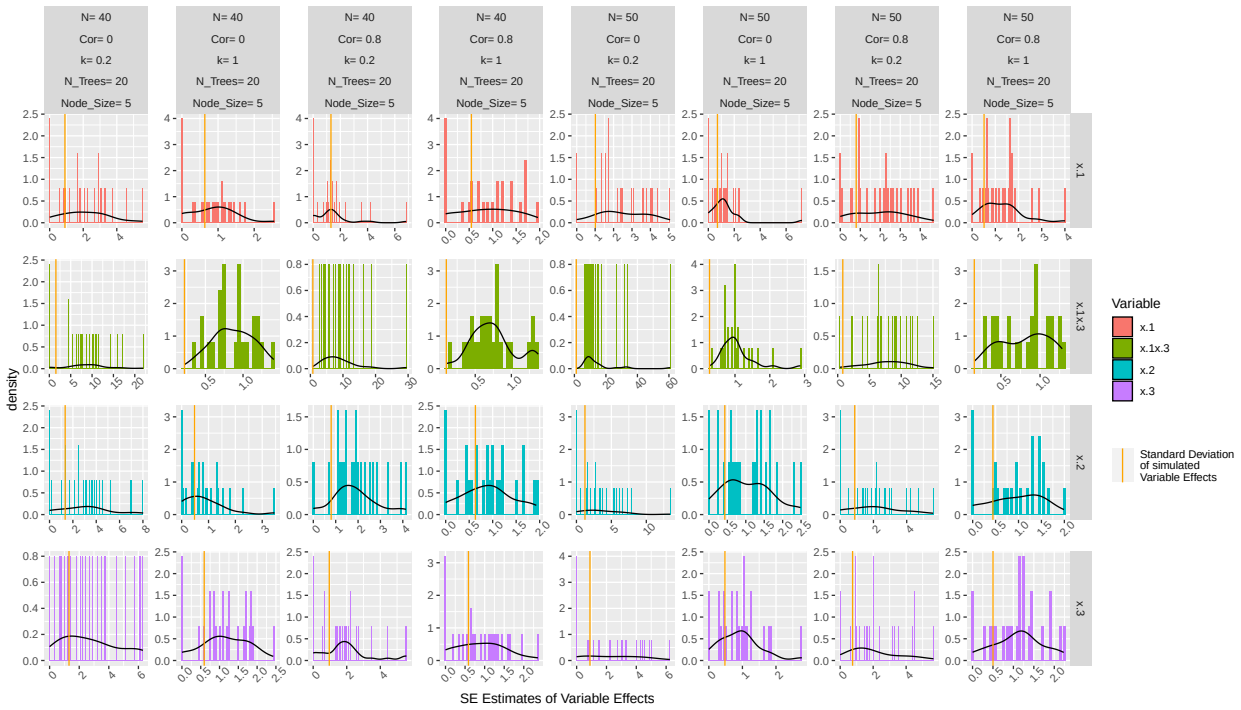
```
se_plot <- plot_se(result)
```

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

```
effect_plots
```


N= 40
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
RF
LM
x.1
x.1x.3
x.2
x.3
Model
Effect Estimates
TrueEffect
Model
LM
RF

```
se_plot
```


N= 40
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 40
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0
k= 1
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 0.2
N_Trees= 20
Node_Size= 5
N= 50
Cor= 0.8
k= 1
N_Trees= 20
Node_Size= 5
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

```
plot_pdps(result)
```