

BE_PCLT_report

Lei Shi

2022-08-17

```
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v tibble  3.1.6      v purrr  0.3.4
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Experiment setup

- $K = 10$, number of factors
- $Q_U = 638$, number of unreplicated arms, each with $N_q = 1$
- $Q_R = 375$, number of small replicated arms, each with $N_q = 2$
- $Q_L = 11$, number of large arms, each with $N_q = 30$.
- $N = 1718$, population

Generate data such that all the k -way ($k \geq 3$) interactions are zero.

Target effects we want to estimate: 'F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3'

Experiment with correct specifications

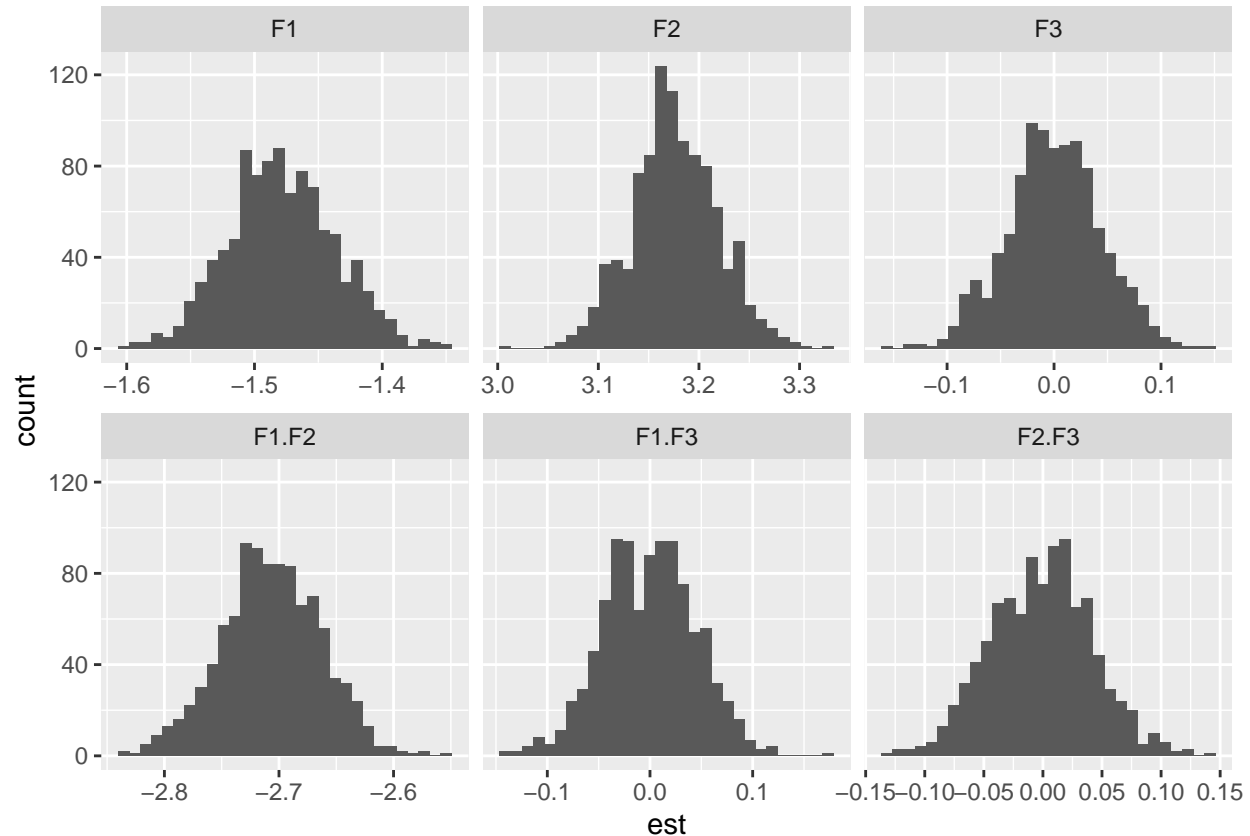
In this experiments, we run 1000 MC trials. The data are generated such that only 'F1', 'F2' and 'F1.F2' have nonzero effects.

```
# report results
record <- readRDS("record_CORRECT_SPEC.RData")
target_effect <- c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3')
# point estimates
```

```
hist_data <- data.frame(
  est = c(t(record[[1]])),
  labs = factor(rep(target_effect, each = 1000), levels = c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3'))
)

ggplot(hist_data, aes(x = est)) +
  facet_wrap(~labs, scales = 'free_x') +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# true effects: -1.481314  3.175201  0.000000 -2.706229  0.000000  0.000000
```

```
# expectation of sd estimator
```

```
target_effect <- c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3')
print(data.frame(
  target_effect = target_effect,
  wls_coverage = diag(apply(sqrt(record[[2]]), MARGIN = c(1,2), sum))/1000,
  ehwc_coverage = diag(apply(sqrt(record[[3]]), MARGIN = c(1,2), sum))/1000,
  lex_coverage = diag(apply(sqrt(record[[4]]), MARGIN = c(1,2), sum))/1000
))
```

```
## Warning in sqrt(record[[3]]): NaNs produced
```

```
## Warning in sqrt(record[[4]]): NaNs produced
```

```
##   target_effect wls_coverage ehwc_coverage lex_coverage
## 1           F1   0.03773547   0.04368083   0.06409644
## 2           F2   0.03773547   0.04368083   0.06409644
```

```
## 3      F3      0.03773547      0.04368083      0.06409644
## 4      F1.F2    0.03773547      0.04368083      0.06409644
## 5      F1.F3    0.03773547      0.04368083      0.06409644
## 6      F2.F3    0.03773547      0.04368083      0.06409644
```

```
# true sd's:
# 0.04368610 0.04368584 0.04368498 0.04368678 0.04368549 0.04368564
```

```
# CI coverage
target_effect <- c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3')
print(data.frame(
  target_effect = target_effect,
  wls_coverage = rowSums(record[[5]])/1000,
  ehw_coverage = rowSums(record[[6]])/1000,
  lex_coverage = rowSums(record[[7]])/1000
))
```

```
##      target_effect wls_coverage ehw_coverage lex_coverage
## 1      F1          0.923        0.958        0.999
## 2      F2          0.921        0.954        0.996
## 3      F3          0.894        0.953        0.994
## 4      F1.F2       0.914        0.955        0.996
## 5      F1.F3       0.909        0.951        0.996
## 6      F2.F3       0.914        0.955        0.998
```

```
# wald inference
print(data.frame(
  method = c('WLS', 'EHW', 'LEX'),
  wald_coverage = c(sum(record[[8]])/1000, sum(record[[9]])/1000, sum(record[[10]])/1000)
))
```

```
##      method wald_coverage
## 1      WLS          0.851
## 2      EHW          0.946
## 3      LEX          1.000
```

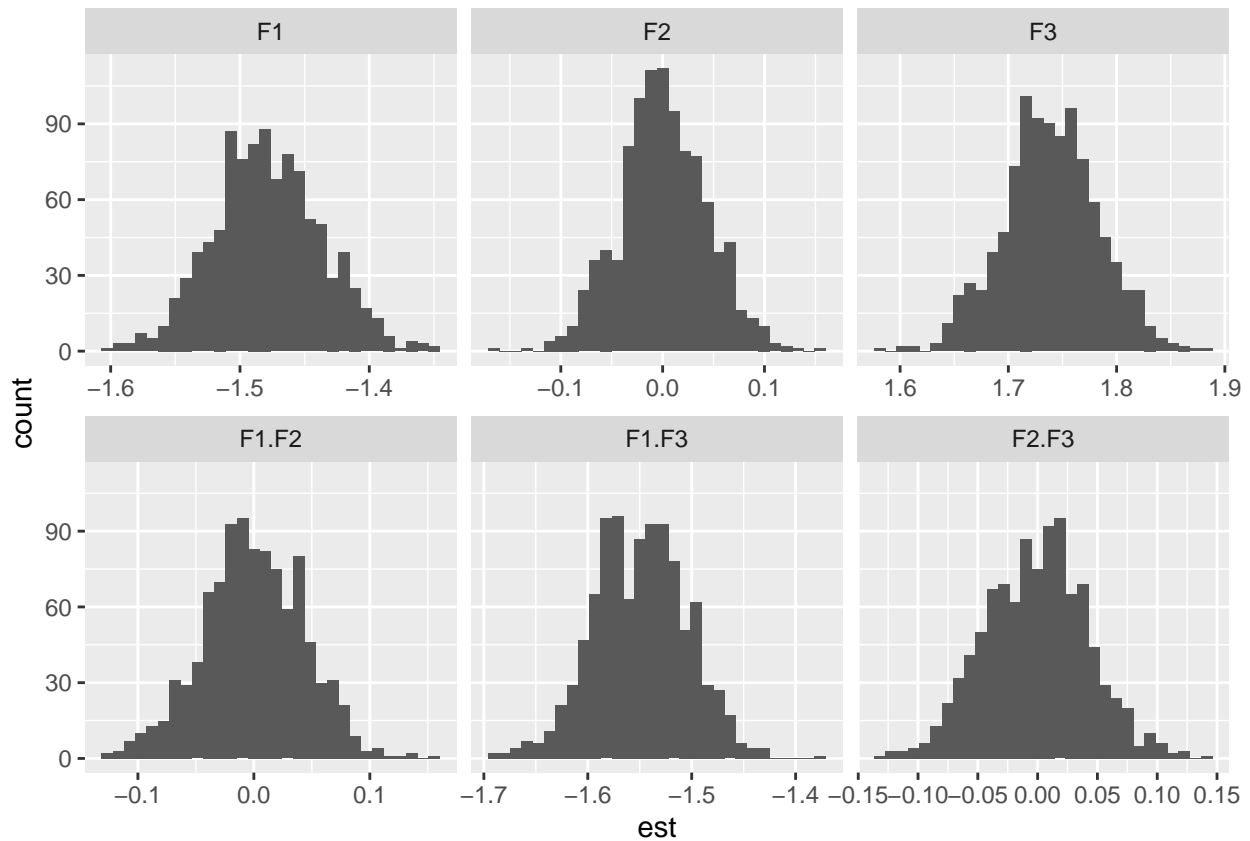
Experiment with incorrect specifications

In this experiments, we run 1000 MC trials. The data are generated such that only 'F1', 'F3', 'F5', 'F7', 'F9' have nonzero main effects. The two-way interactions are generated through strong heredity so some of them are also nonzero.

```
# report results
record <- readRDS("record_INCORRECT_SPEC.RData")
target_effect <- c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3')
# point estimates
hist_data <- data.frame(
  est = c(t(record[[1]])),
  labs = factor(rep(target_effect, each = 1000), levels = c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3'))
)

ggplot(hist_data, aes(x = est)) +
  facet_wrap(~labs, scales = 'free_x') +
  geom_histogram()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# true effects: -1.481314  0.000000  1.738920  0.000000 -1.548822  0.000000
```

```
# expectation of sd estimator
```

```
target_effect <- c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3')
print(data.frame(
  target_effect = target_effect,
  wls_coverage = diag(apply(sqrt(record[[2]]), MARGIN = c(1,2), sum))/1000,
  ehw_coverage = diag(apply(sqrt(record[[3]]), MARGIN = c(1,2), sum))/1000,
  lex_coverage = diag(apply(sqrt(record[[4]]), MARGIN = c(1,2), sum))/1000
))
```

```
## Warning in sqrt(record[[3]]): NaNs produced
```

```
## Warning in sqrt(record[[4]]): NaNs produced
```

```
##   target_effect wls_coverage ehw_coverage lex_coverage
## 1           F1    0.2434383    0.2869593    0.2113441
## 2           F2    0.2434383    0.2869593    0.2113441
## 3           F3    0.2434383    0.2869593    0.2113441
## 4        F1.F2    0.2434383    0.2869593    0.2113441
## 5        F1.F3    0.2434383    0.2869593    0.2113441
## 6        F2.F3    0.2434383    0.2869593    0.2113441
```

```
# true sd's:
```

```
# 0.04368610 0.04368584 0.04368498 0.04368678 0.04368549 0.04368564
```

```
# CI coverage
```

```
target_effect <- c('F1', 'F2', 'F3', 'F1.F2', 'F1.F3', 'F2.F3')
print(data.frame(
  target_effect = target_effect,
```

```

wls_coverage = rowSums(record[[5]])/1000,
ehw_coverage = rowSums(record[[6]])/1000,
lex_coverage = rowSums(record[[7]])/1000
))

##   target_effect wls_coverage ehw_coverage lex_coverage
## 1           F1           1           1           1
## 2           F2           1           1           1
## 3           F3           1           1           1
## 4        F1.F2           1           1           1
## 5        F1.F3           1           1           1
## 6        F2.F3           1           1           1

# wald inference
print(data.frame(
  method = c('WLS', 'EHW', 'LEX'),
  wald_coverage = c(sum(record[[8]])/1000, sum(record[[9]])/1000, sum(record[[10]])/1000)
))

##   method wald_coverage
## 1    WLS           1
## 2    EHW           1
## 3    LEX           1

```