In [1]:
```
# ****************************************
# Variable Selection Simulation Study
#
# Jupyter Notebook Interactive Demonstration!
#
# Emma Tarmey
#
# Started:          03/10/2023
# Most Recent Edit: 13/10/2023
# ****************************************
```

In [2]:
```
# sanity check file location
getwd()
```

'/home/aa22294/Desktop/PhD - Computational Statistics/Projects/Model Selection Sim Study/Code/Jupyter'

In [3]:
```
# pull from R file
source("../R/simulation.R")
#file.show("simulation.R")
```

In [4]:
```
# run simulation
# S = number of scenarios
# N = repetitions for this scenario
# M = number of VS techniques under investigation
# p = number of variables in data (includes id, excludes intercept and ou
# n = synthetic data-set size
run.simulation(S = 4,
               N = 1000,
               M = 6,
               p = 6,
               n = 10000,
               messages = FALSE)
```

```
Scenario  1 / 4

Scenario  2 / 4

Scenario  3 / 4

Scenario  4 / 4
```

In [5]:
```
source("interpret_bias_results.R")

all.results <- get.results.data()

bias.results.s1 <- all.results[[1]]
bias.results.s2 <- all.results[[2]]
bias.results.s3 <- all.results[[3]]
bias.results.s4 <- all.results[[4]]

coef.results.s1 <- all.results[[5]]
coef.results.s2 <- all.results[[6]]
coef.results.s3 <- all.results[[7]]
coef.results.s4 <- all.results[[8]]

all.means <- bias.tables(bias.results.s1, bias.results.s2, bias.results.s
```

```
                        coef.results.s1, coef.results.s2, coef.results.s

s1.bias.means <- all.means[[1]]
s2.bias.means <- all.means[[2]]
s3.bias.means <- all.means[[3]]
s4.bias.means <- all.means[[4]]

s1.bias.means %>% knitr::kable()
s2.bias.means %>% knitr::kable()
s3.bias.means %>% knitr::kable()
s4.bias.means %>% knitr::kable()

bias.plots(s1.bias.means, s2.bias.means, s3.bias.means, s4.bias.means)
```

Raw Bias Values:

Mean Bias of each VS Technique for each Parameter estimate:

Scenario = 1, N = 1000

Scenario = 2, N = 1000

Scenario = 3, N = 1000

Scenario = 4, N = 1000

Technique	Variable	Bias
linear	id	-0.0000011
lasso	id	0.0000000
ridge	id	0.0000941
scad	id	0.0000000
mcp	id	0.0000000
stepwise	id	-0.0000011
linear	c.1	0.0002129
lasso	c.1	0.0000000
ridge	c.1	0.2590366
scad	c.1	-0.0000151
mcp	c.1	-0.0000151
stepwise	c.1	0.0002129
linear	c.2	-0.0002393
lasso	c.2	0.0000000
ridge	c.2	0.1306120
scad	c.2	0.0000060
mcp	c.2	0.0000060
stepwise	c.2	-0.0002393
linear	x.1	0.0001656
lasso	x.1	-0.0314364
ridge	x.1	-0.4254581
scad	x.1	0.0001685
mcp	x.1	0.0001685
stepwise	x.1	0.0001656
linear	x.2	0.0001770
lasso	x.2	0.0303658
ridge	x.2	0.0896503
scad	x.2	0.0001929
mcp	x.2	0.0001929
stepwise	x.2	0.0001770
linear	x.3	0.0002186
lasso	x.3	-0.0322999
ridge	x.3	-0.2443775
scad	x.3	0.0002173
mcp	x.3	0.0002173
stepwise	x.3	0.0002186

Technique	Variable	Bias
linear	id	-0.0000004
lasso	id	0.0000000
ridge	id	0.0000858
scad	id	0.0000000
mcp	id	0.0000000
stepwise	id	-0.0000004
linear	c.1	-0.0007450
lasso	c.1	0.0534689
ridge	c.1	0.3440651
scad	c.1	0.0000000
mcp	c.1	0.0000000
stepwise	c.1	-0.0007450
linear	c.2	0.0002019
lasso	c.2	0.0000009
ridge	c.2	0.1197756
scad	c.2	0.0000091
mcp	c.2	0.0000091
stepwise	c.2	0.0002019
linear	x.1	-0.0000369
lasso	x.1	-0.0199894
ridge	x.1	-0.2809589
scad	x.1	-0.0000382
mcp	x.1	-0.0000382
stepwise	x.1	-0.0000369
linear	x.2	0.0008361
lasso	x.2	-0.1022298
ridge	x.2	-0.3009074
scad	x.2	-0.0000748
mcp	x.2	-0.0000748
stepwise	x.2	0.0008361
linear	x.3	-0.0003884
lasso	x.3	0.0021412
ridge	x.3	-0.1175053
scad	x.3	-0.0003904
mcp	x.3	-0.0003904
stepwise	x.3	-0.0003884

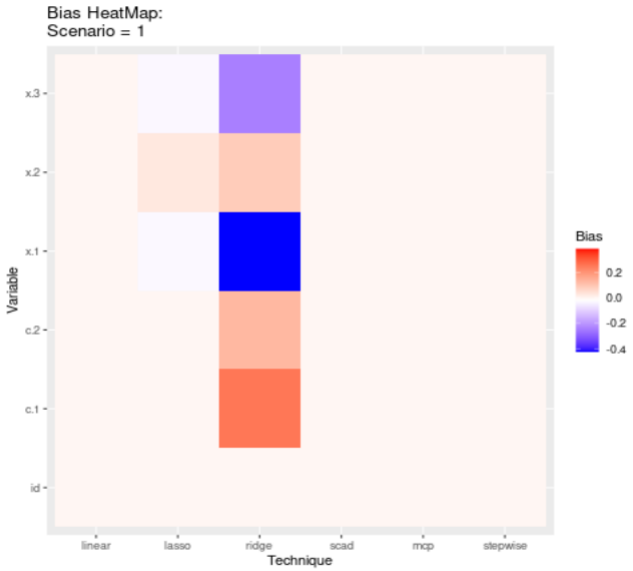
Technique	Variable	Bias
linear	id	0.0000000
lasso	id	0.0000000
ridge	id	0.0000655
scad	id	0.0000000
mcp	id	0.0000000
stepwise	id	0.0000000
linear	c.1	-0.0001543
lasso	c.1	0.0172602
ridge	c.1	0.2162824
scad	c.1	0.0000000
mcp	c.1	0.0000000
stepwise	c.1	-0.0001543
linear	c.2	-0.0001884
lasso	c.2	0.0000000
ridge	c.2	0.0898231
scad	c.2	0.0000000
mcp	c.2	0.0000000
stepwise	c.2	-0.0001884
linear	x.1	-0.0001220
lasso	x.1	-0.0746999
ridge	x.1	-0.2718800
scad	x.1	-0.0002468
mcp	x.1	-0.0002468
stepwise	x.1	-0.0001220
linear	x.2	0.0001174
lasso	x.2	-0.0915269
ridge	x.2	-0.2865494
scad	x.2	0.0000803
mcp	x.2	0.0000803
stepwise	x.2	0.0001174
linear	x.3	-0.0000996
lasso	x.3	-0.0024613
ridge	x.3	-0.1697767
scad	x.3	-0.0001001
mcp	x.3	-0.0001001
stepwise	x.3	-0.0000996

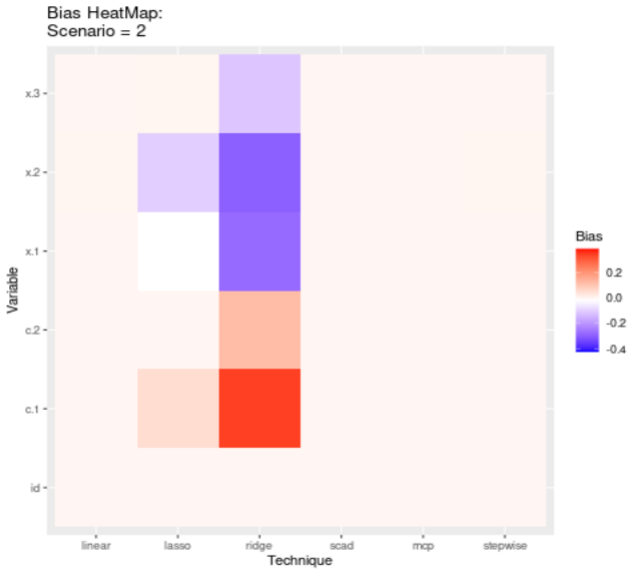
```
|Technique |Variable |       Bias|
|:---------|:--------|----------:|
|linear    |id       |  0.0000000|
|lasso     |id       |  0.0000000|
|ridge     |id       |  0.0000607|
|scad      |id       |  0.0000000|
|mcp       |id       |  0.0000000|
|stepwise  |id       |  0.0000000|
|linear    |c.1      | -0.0002936|
|lasso     |c.1      |  0.0091904|
|ridge     |c.1      |  0.1449353|
|scad      |c.1      |  0.0000000|
|mcp       |c.1      |  0.0000000|
|stepwise  |c.1      | -0.0002936|
|linear    |c.2      |  0.0000298|
|lasso     |c.2      |  0.0000000|
|ridge     |c.2      |  0.1036923|
|scad      |c.2      |  0.0000000|
|mcp       |c.2      |  0.0000000|
|stepwise  |c.2      |  0.0000298|
|linear    |x.1      |  0.0012182|
|lasso     |x.1      | -0.0226102|
|ridge     |x.1      |  0.3871795|
|scad      |x.1      | -0.0000205|
|mcp       |x.1      | -0.0000205|
|stepwise  |x.1      |  0.0012182|
|linear    |x.2      | -0.0002586|
|lasso     |x.2      | -0.1140752|
|ridge     |x.2      | -0.2859044|
|scad      |x.2      | -0.0002924|
|mcp       |x.2      | -0.0002924|
|stepwise  |x.2      | -0.0002586|
|linear    |x.3      | -0.0001294|
|lasso     |x.3      |  0.0015854|
|ridge     |x.3      | -0.1799660|
|scad      |x.3      | -0.0001336|
|mcp       |x.3      | -0.0001336|
|stepwise  |x.3      | -0.0001294|
[1] 36
[1] 3
[1] NA


Changing plot `p1`

Changing plot `p2`

Changing plot `p3`

Changing plot `p4`

[1] "/home/aa22294/Desktop/PhD - Computational Statistics/Projects/Model S
election Sim Study/Code/R"
```

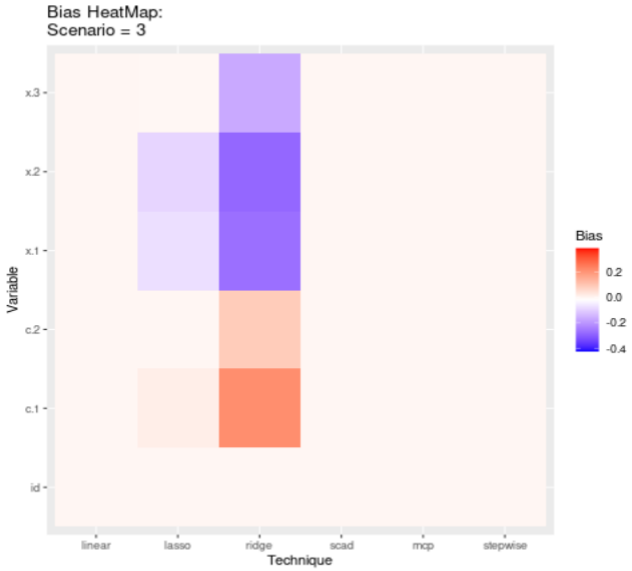
png: 2

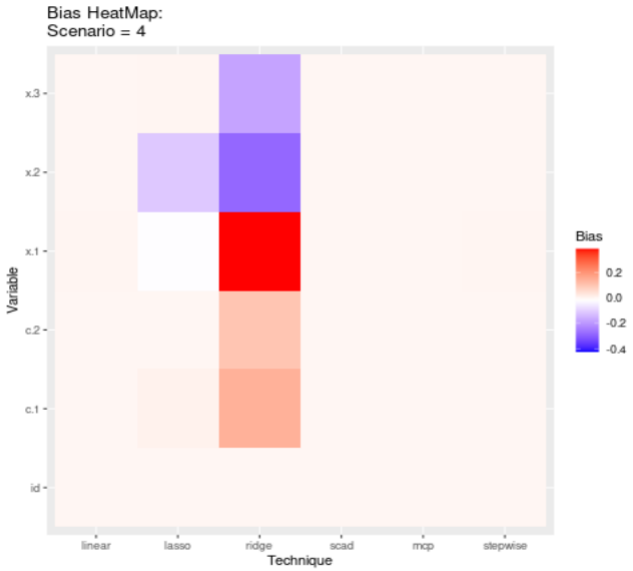
In [6]:
```
library("png")

plot.new()
pp <- readPNG("../plots/bias_s1.png")
```

```
rasterImage(pp, 0.00, 0.00, 1.00, 1.00)

plot.new()
pp <- readPNG("../plots/bias_s2.png")
rasterImage(pp, 0.00, 0.00, 1.00, 1.00)

plot.new()
pp <- readPNG("../plots/bias_s3.png")
rasterImage(pp, 0.00, 0.00, 1.00, 1.00)

plot.new()
pp <- readPNG("../plots/bias_s4.png")
rasterImage(pp, 0.00, 0.00, 1.00, 1.00)
```




Bias HeatMap:
Scenario = 2
x.3
x.2
x.1
c.2
c.1
id
Variable
linear
lasso
ridge
scad
mcp
stepwise
Technique
Bias
0.2
0.0
-0.2
-0.4


Bias HeatMap:
Scenario = 3
x.3
x.2
x.1
c.2
c.1
id
Variable
linear
lasso
ridge
scad
mcp
stepwise
Technique
Bias
0.2
0.0
-0.2
-0.4



In [8]:
```
source("interpret_coef_results.R")

all.results <- get.results.data()

bias.results.s1 <- all.results[[1]]
bias.results.s2 <- all.results[[2]]
bias.results.s3 <- all.results[[3]]
bias.results.s4 <- all.results[[4]]

coef.results.s1 <- all.results[[5]]
coef.results.s2 <- all.results[[6]]
coef.results.s3 <- all.results[[7]]
coef.results.s4 <- all.results[[8]]


lr.coef <- coef.tables(method = "linear", coef.results.s1, coef.results.s

lr.coef.summary.s1 <- lr.coef[[1]]
lr.coef.summary.s2 <- lr.coef[[2]]
lr.coef.summary.s3 <- lr.coef[[3]]
lr.coef.summary.s4 <- lr.coef[[4]]

lr.coef.summary.s1 %>% knitr::kable()
lr.coef.summary.s2 %>% knitr::kable()
lr.coef.summary.s3 %>% knitr::kable()
lr.coef.summary.s4 %>% knitr::kable()


lasso.coef <- coef.tables(method = "lasso", coef.results.s1, coef.results

lasso.coef.summary.s1 <- lasso.coef[[1]]
lasso.coef.summary.s2 <- lasso.coef[[2]]
lasso.coef.summary.s3 <- lasso.coef[[3]]
lasso.coef.summary.s4 <- lasso.coef[[4]]
```

```
lasso.coef.summary.s1 %>% knitr::kable()
lasso.coef.summary.s2 %>% knitr::kable()
lasso.coef.summary.s3 %>% knitr::kable()
lasso.coef.summary.s4 %>% knitr::kable()


ridge.coef <- coef.tables(method = "ridge", coef.results.s1, coef.results

ridge.coef.summary.s1 <- ridge.coef[[1]]
ridge.coef.summary.s2 <- ridge.coef[[2]]
ridge.coef.summary.s3 <- ridge.coef[[3]]
ridge.coef.summary.s4 <- ridge.coef[[4]]

ridge.coef.summary.s1 %>% knitr::kable()
ridge.coef.summary.s2 %>% knitr::kable()
ridge.coef.summary.s3 %>% knitr::kable()
ridge.coef.summary.s4 %>% knitr::kable()


scad.coef <- coef.tables(method = "scad", coef.results.s1, coef.results.s

scad.coef.summary.s1 <- scad.coef[[1]]
scad.coef.summary.s2 <- scad.coef[[2]]
scad.coef.summary.s3 <- scad.coef[[3]]
scad.coef.summary.s4 <- scad.coef[[4]]

scad.coef.summary.s1 %>% knitr::kable()
scad.coef.summary.s2 %>% knitr::kable()
scad.coef.summary.s3 %>% knitr::kable()
scad.coef.summary.s4 %>% knitr::kable()


mcp.coef <- coef.tables(method = "mcp", coef.results.s1, coef.results.s2,

mcp.coef.summary.s1 <- mcp.coef[[1]]
mcp.coef.summary.s2 <- mcp.coef[[2]]
mcp.coef.summary.s3 <- mcp.coef[[3]]
mcp.coef.summary.s4 <- mcp.coef[[4]]

mcp.coef.summary.s1 %>% knitr::kable()
mcp.coef.summary.s2 %>% knitr::kable()
mcp.coef.summary.s3 %>% knitr::kable()
mcp.coef.summary.s4 %>% knitr::kable()


step.coef <- coef.tables(method = "stepwise", coef.results.s1, coef.resul

step.coef.summary.s1 <- step.coef[[1]]
step.coef.summary.s2 <- step.coef[[2]]
step.coef.summary.s3 <- step.coef[[3]]
step.coef.summary.s4 <- step.coef[[4]]

step.coef.summary.s1 %>% knitr::kable()
step.coef.summary.s2 %>% knitr::kable()
step.coef.summary.s3 %>% knitr::kable()
step.coef.summary.s4 %>% knitr::kable()
```

linear Parameter Estimates for each Scenario

Variable	True	Mean	SD
id	0	-0.0000011	0.0000100
c.1	0	0.0002129	0.0117545
c.2	0	-0.0002393	0.0117414
x.1	1	1.0001656	0.0040014
x.2	1	1.0001770	0.0121616
x.3	1	1.0002186	0.0085538

Variable	True	Mean	SD
id	0	-0.0000004	0.0000106
c.1	0	-0.0007450	0.0195969
c.2	0	0.0002019	0.0121621
x.1	1	0.9999631	0.0040065
x.2	1	1.0008361	0.0301511
x.3	1	0.9996116	0.0082889

Variable	True	Mean	SD
id	0	0.0000000	0.0000036
c.1	0	-0.0001543	0.0042633
c.2	0	-0.0001884	0.0040704
x.1	1	0.9998780	0.0096482
x.2	1	1.0001174	0.0096590
x.3	1	0.9999004	0.0029243

Variable	True	Mean	SD
id	0	0.0000000	0.0000034
c.1	0	-0.0002936	0.0055875
c.2	0	0.0000298	0.0040775
x.1	1	1.0012182	0.0345646
x.2	1	0.9997414	0.0099956
x.3	1	0.9998706	0.0029080

lasso Parameter Estimates for each Scenario

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0000000	0.0000000
c.2	0	0.0000000	0.0000000
x.1	1	0.9685636	0.0040449
x.2	1	1.0303658	0.0057289
x.3	1	0.9677001	0.0075259

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0534689	0.0178708
c.2	0	0.0000009	0.0000174
x.1	1	0.9800106	0.0040051
x.2	1	0.8977702	0.0302913
x.3	1	1.0021412	0.0059855

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0172602	0.0036009
c.2	0	0.0000000	0.0000000
x.1	1	0.9253001	0.0096856
x.2	1	0.9084731	0.0097729
x.3	1	0.9975387	0.0021166

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0091904	0.0052597
c.2	0	0.0000000	0.0000000
x.1	1	0.9773898	0.0334488
x.2	1	0.8859248	0.0099161
x.3	1	1.0015854	0.0021478

ridge Parameter Estimates for each Scenario

Variable	True	Mean	SD
id	0	0.0000941	0.0000086
c.1	0	0.2590366	0.0100189
c.2	0	0.1306120	0.0099282
x.1	1	0.5745419	0.0031405
x.2	1	1.0896503	0.0074368
x.3	1	0.7556225	0.0077849

Variable	True	Mean	SD
id	0	0.0000858	0.0000086
c.1	0	0.3440651	0.0093606
c.2	0	0.1197756	0.0106119
x.1	1	0.7190411	0.0034303
x.2	1	0.6990926	0.0158523
x.3	1	0.8824947	0.0061420

Variable	True	Mean	SD
id	0	0.0000655	0.0000031
c.1	0	0.2162824	0.0036935
c.2	0	0.0898231	0.0035476
x.1	1	0.7281200	0.0089199
x.2	1	0.7134506	0.0093327
x.3	1	0.8302233	0.0022635

Variable	True	Mean	SD
id	0	0.0000607	0.0000029
c.1	0	0.1449353	0.0036031
c.2	0	0.1036923	0.0036629
x.1	1	1.3871795	0.0235988
x.2	1	0.7140956	0.0093318
x.3	1	0.8200340	0.0023095

scad Parameter Estimates for each Scenario

Variable	True	Mean	SD
id	0	0.0000000	0.0000008
c.1	0	-0.0000151	0.0006567
c.2	0	0.0000060	0.0004741
x.1	1	1.0001685	0.0040009
x.2	1	1.0001929	0.0121614
x.3	1	1.0002173	0.0085505

Variable	True	Mean	SD
id	0	0.0000000	0.0000004
c.1	0	0.0000000	0.0000011
c.2	0	0.0000091	0.0005234
x.1	1	0.9999618	0.0040021
x.2	1	0.9999252	0.0186695
x.3	1	0.9996096	0.0082867

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0000000	0.0000000
c.2	0	0.0000000	0.0000000
x.1	1	0.9997532	0.0091046
x.2	1	1.0000803	0.0095684
x.3	1	0.9998999	0.0029246

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0000000	0.0000000
c.2	0	0.0000000	0.0000000
x.1	1	0.9999795	0.0259779
x.2	1	0.9997076	0.0096593
x.3	1	0.9998664	0.0029090

mcp Parameter Estimates for each Scenario

Variable	True	Mean	SD
id	0	0.0000000	0.0000008
c.1	0	-0.0000151	0.0006567
c.2	0	0.0000060	0.0004741
x.1	1	1.0001685	0.0040009
x.2	1	1.0001929	0.0121614
x.3	1	1.0002173	0.0085505

Variable	True	Mean	SD
id	0	0.0000000	0.0000004
c.1	0	0.0000000	0.0000011
c.2	0	0.0000091	0.0005234
x.1	1	0.9999618	0.0040021
x.2	1	0.9999252	0.0186695
x.3	1	0.9996096	0.0082867

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0000000	0.0000000
c.2	0	0.0000000	0.0000000
x.1	1	0.9997532	0.0091046
x.2	1	1.0000803	0.0095684
x.3	1	0.9998999	0.0029246

Variable	True	Mean	SD
id	0	0.0000000	0.0000000
c.1	0	0.0000000	0.0000000
c.2	0	0.0000000	0.0000000
x.1	1	0.9999795	0.0259779
x.2	1	0.9997076	0.0096593
x.3	1	0.9998664	0.0029090

stepwise Parameter Estimates for each Scenario

Variable	True	Mean	SD
id	0	-0.0000011	0.0000100
c.1	0	0.0002129	0.0117545
c.2	0	-0.0002393	0.0117414
x.1	1	1.0001656	0.0040014
x.2	1	1.0001770	0.0121616
x.3	1	1.0002186	0.0085538

Variable	True	Mean	SD
id	0	-0.0000004	0.0000106
c.1	0	-0.0007450	0.0195969
c.2	0	0.0002019	0.0121621
x.1	1	0.9999631	0.0040065
x.2	1	1.0008361	0.0301511
x.3	1	0.9996116	0.0082889

Variable	True	Mean	SD
id	0	0.0000000	0.0000036
c.1	0	-0.0001543	0.0042633
c.2	0	-0.0001884	0.0040704
x.1	1	0.9998780	0.0096482
x.2	1	1.0001174	0.0096590
x.3	1	0.9999004	0.0029243

Variable	True	Mean	SD
id	0	0.0000000	0.0000034
c.1	0	-0.0002936	0.0055875
c.2	0	0.0000298	0.0040775
x.1	1	1.0012182	0.0345646
x.2	1	0.9997414	0.0099956
x.3	1	0.9998706	0.0029080

In [ ]: