# BIOS 731: HW2

Randy Parrish

2026-02-11

---

## Problem 0

GitHub Repo: https://github.com/rndparr/bios731_hw2_parrish

---

## Problem 1.1 ADEMP Structure

- **A (Aim):** The goal is to evaluate three methods for constructing a 95% confidence interval for $\hat{\beta}_{treatment}$.
- **D (Data-generating mechanism):** are generated from a normal distribution. Errors are generated from either a normal or $t$ distribution, depending on scenario. The sample size, true treatment effect, and error distributions vary across scenarios.
- **E (Estimand):** We are most interested in $\beta_{treatment}$.
- **M (Methods):** Wald CI, Nonparametric bootstrap percentile intervals, and Nonparametric bootstrap $t$ intervals
- **P (Performance measures):** The performance measures are Bias($\hat{\beta}$), probability of coverage of the 95% CI, standard error behavior, and computation time.
- **Number of simulation scenarios:** There are 18 scenarios: 3 sample sizes x 3 true treatment effects x 2 error distributions=18 scenarios.

---

## Problem 1.2 nSim

Based on a desired coverage of 95% with Monte Carlo error of no more than 1%,

$$n_{sim} = \frac{co\hat{v}er(1 - co\hat{v}er)}{(SE(co\hat{v}er))^2} = 475$$

---

## Problem 1.3 Implementation

The main script for conducting simulations is `./source/run_sim_i.R`. It requires 2 arguments: `i`, the scenario number; and `ncores`, the number of cores to use for parallel processing. It uses functions from these scripts located in `./source/`: `utility.R`, `simulate_data.R`, `ci_coverage.R`, `models.R`, `error_funcs.R`

Due to the computational complexity, the simulation study was run on the RHPC cluster. The jobs were submitted to the cluster by submitting the `./source/sbatch_run_sim_i.sh` script to the cluster as an array job. The `doFuture`,`doRNG`, and `foreach` packages were used to parallelize jobs.

Once the jobs were finished, I transferred the intermmediary data files for each scenario to my local computer and used `./source/merge_data.R` to combine the data contained in each into a single dataframe, saved as `./data/all_data.Rds`. An example of bash commands that can be used to submit such a job is shown in the `README.md` file.

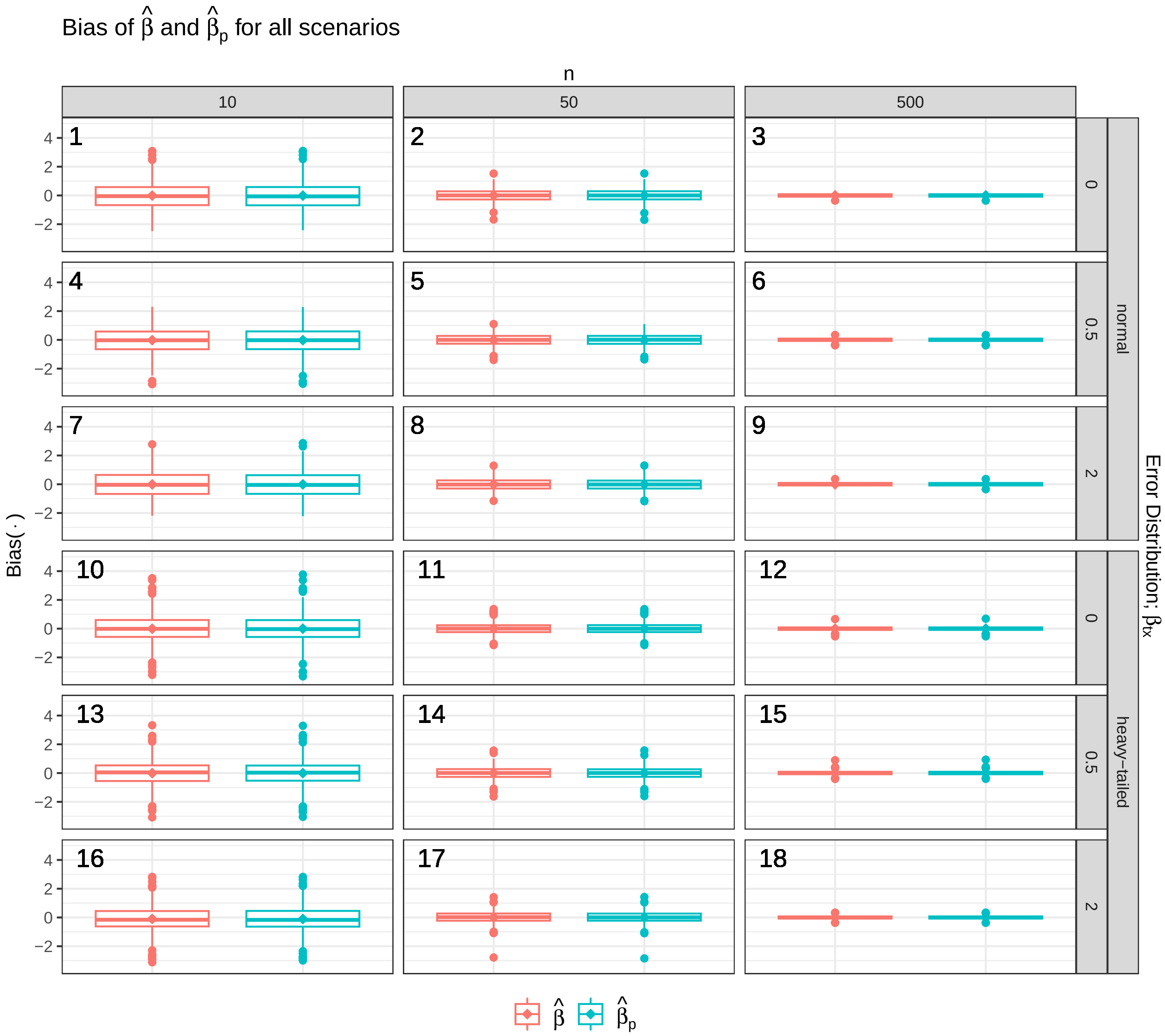
The `./analysis/HW2_final_report.Rmd` file used to generate this pdf sources `./source/tables.R` and `./source/plot_functions.R`. Plotting libraries are loaded in the latter file, which also creates a melted data frame, `mdat`, for use in some plots.

---

## Problem 1.4 Results summary

**Bias($\hat{\beta}$)**

| Scenario | n | $\beta_{tx}$ | Error Distribution | Mean Bias($\hat{\beta}$) | Mean Bias($\hat{\beta}_{boot}$) |
|---:|---:|---:|---|---:|---:|
| 1 | 10 | 0.0 | normal | -0.0086 | -0.0120 |
| 2 | 50 | 0.0 | normal | 0.0063 | 0.0065 |
| 3 | 500 | 0.0 | normal | 0.0029 | 0.0030 |
| 4 | 10 | 0.5 | normal | -0.0220 | -0.0219 |
| 5 | 50 | 0.5 | normal | -0.0173 | -0.0158 |
| 6 | 500 | 0.5 | normal | 0.0034 | 0.0030 |
| 7 | 10 | 2.0 | normal | -0.0099 | -0.0090 |
| 8 | 50 | 2.0 | normal | -0.0128 | -0.0132 |
| 9 | 500 | 2.0 | normal | 0.0017 | 0.0019 |
| 10 | 10 | 0.0 | heavy-tailed | -0.0231 | -0.0202 |
| 11 | 50 | 0.0 | heavy-tailed | 0.0043 | 0.0047 |
| 12 | 500 | 0.0 | heavy-tailed | -0.0007 | -0.0010 |
| 13 | 10 | 0.5 | heavy-tailed | -0.0161 | -0.0157 |
| 14 | 50 | 0.5 | heavy-tailed | 0.0033 | 0.0029 |
| 15 | 500 | 0.5 | heavy-tailed | 0.0037 | 0.0034 |
| 16 | 10 | 2.0 | heavy-tailed | -0.0776 | -0.0816 |
| 17 | 50 | 2.0 | heavy-tailed | 0.0127 | 0.0118 |
| 18 | 500 | 2.0 | heavy-tailed | 0.0018 | 0.0016 |


Bias of $\hat{\beta}$ and $\hat{\beta}_p$ for all scenarios
n
10
50
500
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
4
2
0
−2
Bias( · )
0
0.5
2
normal
heavy−tailed
Error Distribution; $\beta_{tx}$
$\hat{\beta}$
$\hat{\beta}_p$

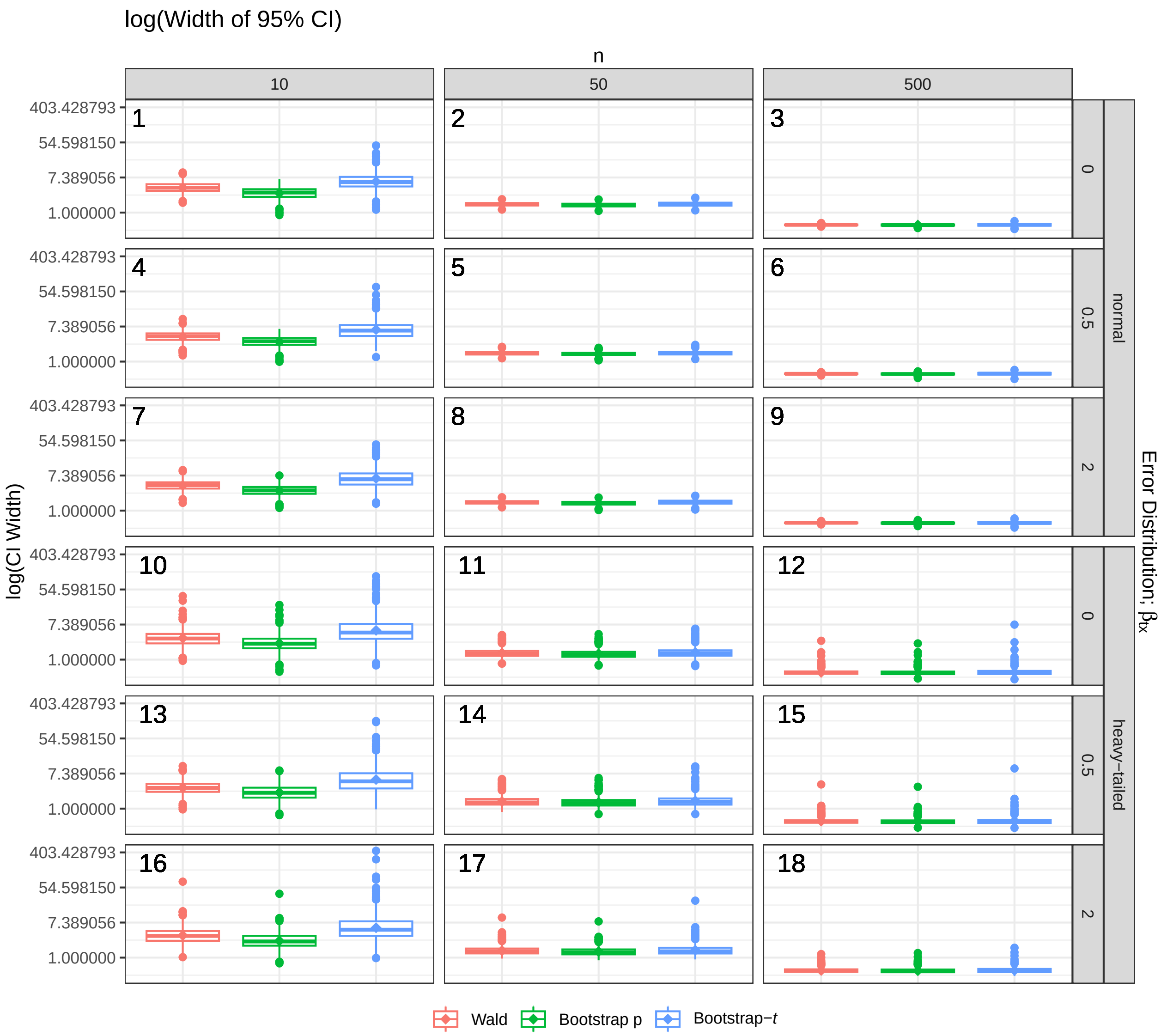
**Coverage of the 95% CI for $\hat{\beta}$**

The following table shows, for all scenarios, the proportion of CIs which covered the true beta value for all models.

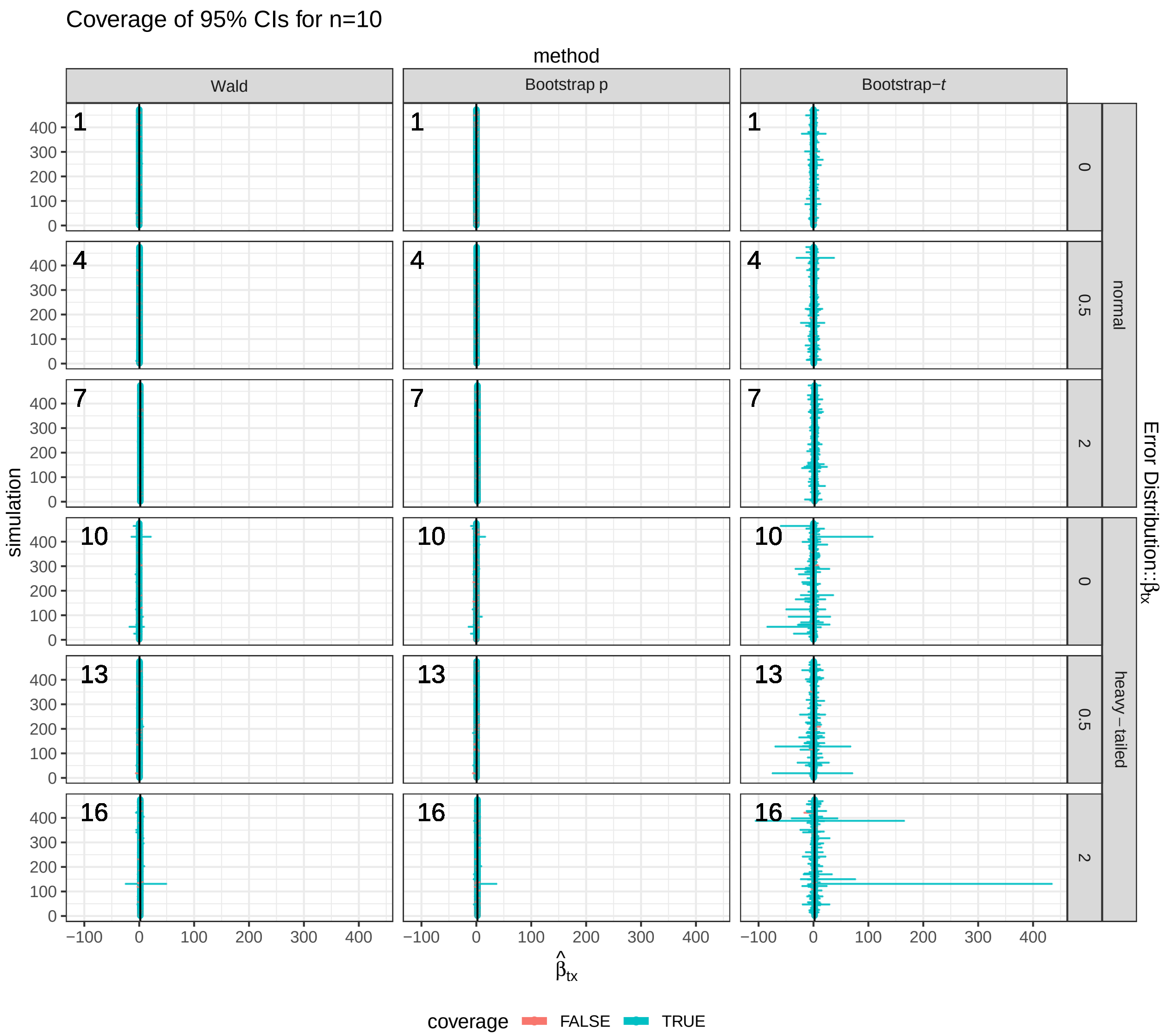
| | | | | Proportion of 95% CIs containing $\beta_{tx}$ | | |
|---|---|---|---|---|---|---|
| Scenario | n | $\beta_{tx}$ | Error Distribution | Wald CI | Boot p CI | Bootstrap-$t$ CI |
| 1 | 10 | 0.0 | normal | 0.9326 | 0.8653 | 0.9663 |
| 2 | 50 | 0.0 | normal | 0.9284 | 0.9200 | 0.9200 |
| 3 | 500 | 0.0 | normal | 0.9579 | 0.9558 | 0.9600 |
| 4 | 10 | 0.5 | normal | 0.9495 | 0.8779 | 0.9684 |
| 5 | 50 | 0.5 | normal | 0.9474 | 0.9305 | 0.9389 |
| 6 | 500 | 0.5 | normal | 0.9537 | 0.9537 | 0.9621 |
| 7 | 10 | 2.0 | normal | 0.9663 | 0.8884 | 0.9726 |
| 8 | 50 | 2.0 | normal | 0.9474 | 0.9389 | 0.9516 |
| 9 | 500 | 2.0 | normal | 0.9453 | 0.9432 | 0.9516 |
| 10 | 10 | 0.0 | heavy-tailed | 0.9558 | 0.8295 | 0.9432 |
| 11 | 50 | 0.0 | heavy-tailed | 0.9558 | 0.9305 | 0.9242 |
| 12 | 500 | 0.0 | heavy-tailed | 0.9579 | 0.9411 | 0.9326 |
| 13 | 10 | 0.5 | heavy-tailed | 0.9516 | 0.8547 | 0.9726 |
| 14 | 50 | 0.5 | heavy-tailed | 0.9516 | 0.9242 | 0.9200 |
| 15 | 500 | 0.5 | heavy-tailed | 0.9326 | 0.9179 | 0.9137 |
| 16 | 10 | 2.0 | heavy-tailed | 0.9621 | 0.8800 | 0.9558 |
| 17 | 50 | 2.0 | heavy-tailed | 0.9663 | 0.9411 | 0.9474 |
| 18 | 500 | 2.0 | heavy-tailed | 0.9663 | 0.9600 | 0.9495 |

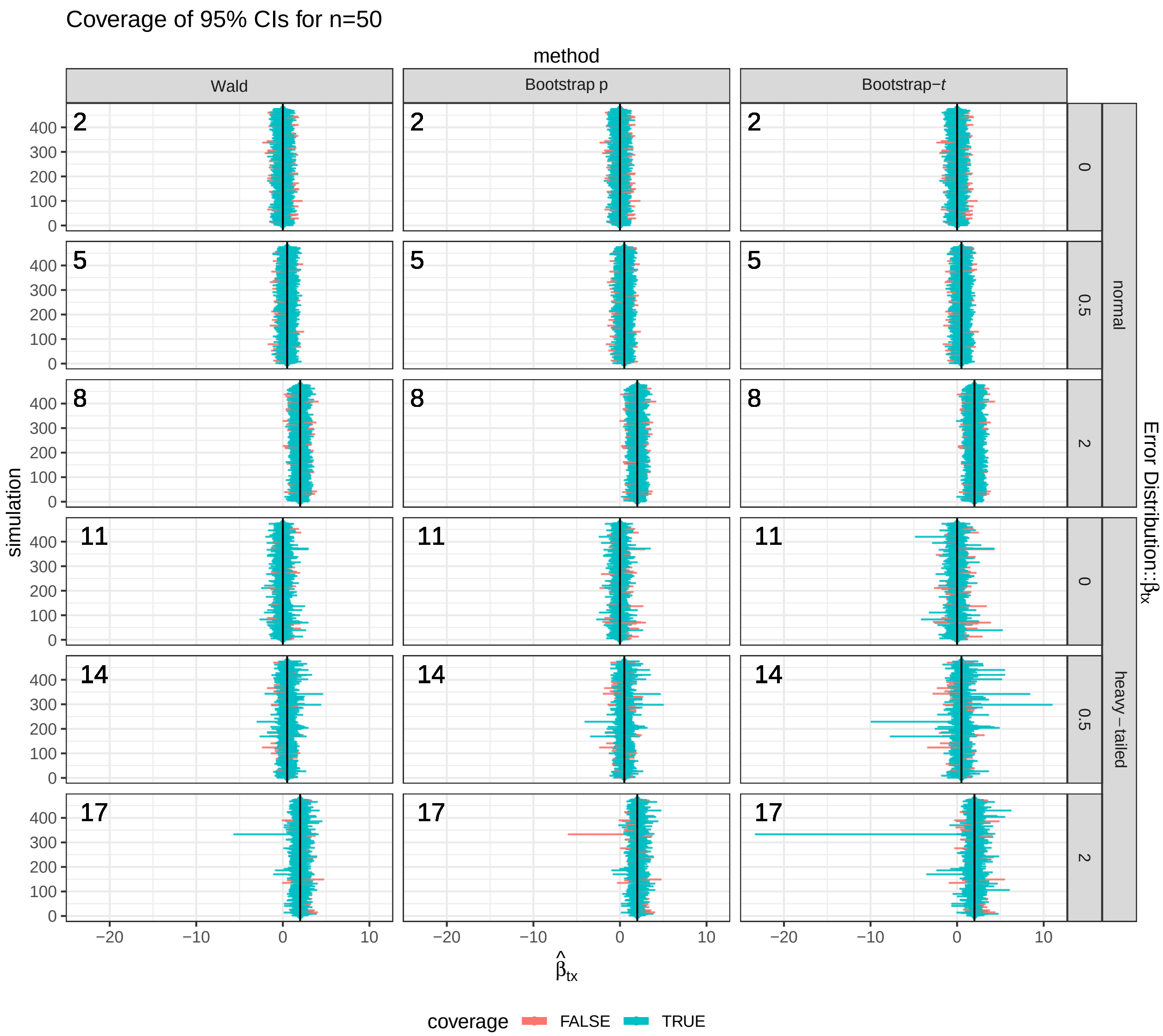
The coverage for all scenarios is relatively close to the nominal coverage of 95%, so the proportion of simulations where coverage was NOT achieved was plotted in order to better visualize the differences between methods under different scenarios. The distribution of 95% CI width were also plotted. The actual confidence intervals for all simulations, grouped by $n$, are also given below.


Proportion of simulations where the 95% CI does NOT include β_tx
n
10
50
500
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
0
0.5
2
normal
heavy−tailed
Proportion
Error Distribution; β_tx
Wald
Bootstrap p
Bootstrap−t
dashed line at α=0.05

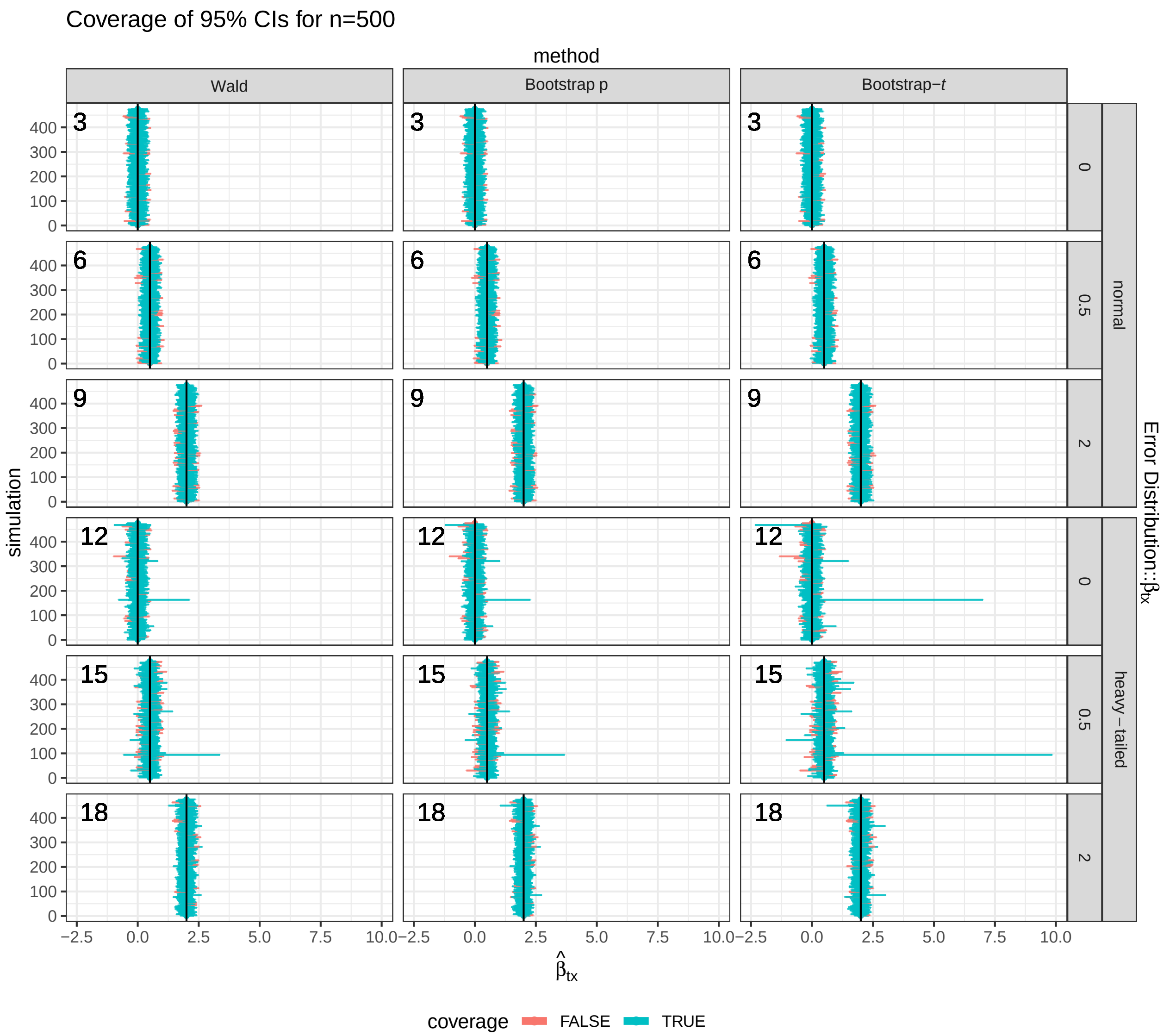


log(Width of 95% CI)
n
10
50
500
log(CI Width)
403.428793
54.598150
7.389056
1.000000
Error Distribution; $\beta_{tx}$
normal
heavy−tailed
0
0.5
2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Wald
Bootstrap p
Bootstrap−*t*

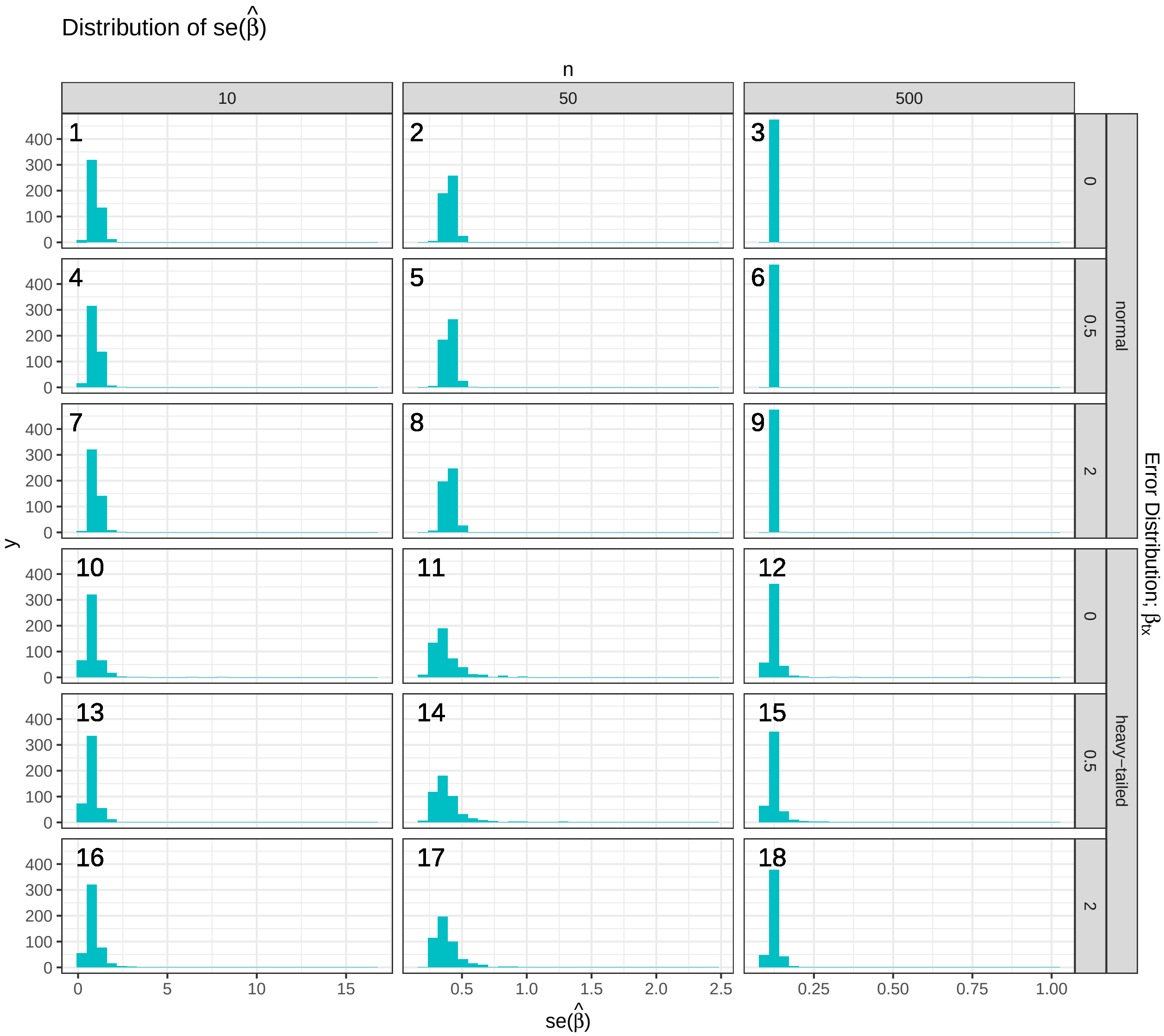

Coverage of 95% CIs for n=10
method
Wald
Bootstrap p
Bootstrap−t
simulation
Error Distribution::βtx
normal
heavy−tailed
0
0.5
2
1
4
7
10
13
16
400
300
200
100
0
−100
0
100
200
300
400
β̂tx
coverage
FALSE
TRUE



Coverage of 95% CIs for n=50
method
Wald
Bootstrap p
Bootstrap−t
normal
heavy−tailed
Error Distribution::$\beta_{tx}$
simulation
$\hat{\beta}_{tx}$
coverage
FALSE
TRUE


Coverage of 95% CIs for n=500
method
Wald
Bootstrap p
Bootstrap−t
simulation
Error Distribution::$\beta_{tx}$
normal
heavy−tailed
0
0.5
2
3
6
9
12
15
18
$\hat{\beta}_{tx}$
coverage
FALSE
TRUE

**Distribution of $se(\hat{\beta})$**

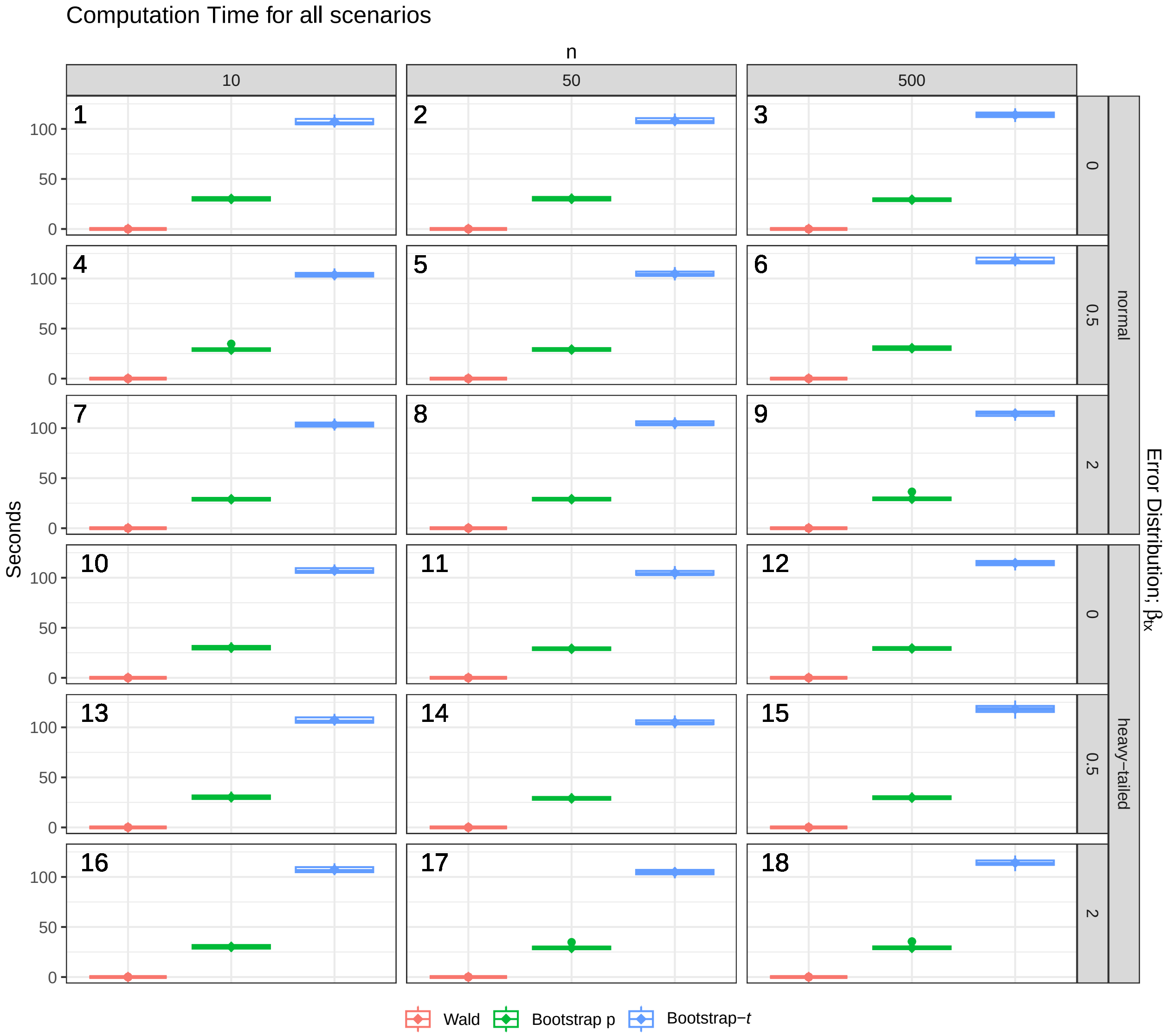



**Computation Time**

In the structure of the code, both the Bootstrap p and Bootstrap-$t$ are run within the same function call so that. Computation time for the combined bootstrap function is calculated and Bootstrap-$t$ is calculated within the function. The Bootstrap p time is assumed to be the combined time minus the Bootstrap-$t$ time.

Based on this table, the following plot was not faceted by scenario parameters like previous plots.

| | | | | Average Time (sec) | | |
|---|---|---|---|---|---|---|
| Scenario | n | $\beta_{tx}$ | Error Distribution | Wald | Bootstrap p | Bootstrap-$t$ |
| 1 | 10 | 0.0 | normal | 0.0022 | 30.0783 | 107.1470 |
| 2 | 50 | 0.0 | normal | 0.0022 | 30.1941 | 108.1709 |
| 3 | 500 | 0.0 | normal | 0.0023 | 29.2366 | 113.9620 |
| 4 | 10 | 0.5 | normal | 0.0021 | 28.9245 | 103.6387 |
| 5 | 50 | 0.5 | normal | 0.0021 | 29.0033 | 104.6538 |
| 6 | 500 | 0.5 | normal | 0.0023 | 30.3679 | 117.9361 |
| 7 | 10 | 2.0 | normal | 0.0021 | 28.9136 | 103.4812 |
| 8 | 50 | 2.0 | normal | 0.0021 | 28.9775 | 104.6675 |
| 9 | 500 | 2.0 | normal | 0.0023 | 29.4220 | 114.3792 |
| 10 | 10 | 0.0 | heavy-tailed | 0.0021 | 30.0824 | 107.1756 |
| 11 | 50 | 0.0 | heavy-tailed | 0.0021 | 28.9826 | 104.6875 |
| 12 | 500 | 0.0 | heavy-tailed | 0.0022 | 29.3139 | 114.4783 |
| 13 | 10 | 0.5 | heavy-tailed | 0.0022 | 30.1607 | 107.1935 |
| 14 | 50 | 0.5 | heavy-tailed | 0.0021 | 28.9776 | 104.8511 |
| 15 | 500 | 0.5 | heavy-tailed | 0.0024 | 29.6787 | 118.2160 |
| 16 | 10 | 2.0 | heavy-tailed | 0.0022 | 30.1399 | 107.2286 |
| 17 | 50 | 2.0 | heavy-tailed | 0.0021 | 29.0942 | 104.7591 |
| 18 | 500 | 2.0 | heavy-tailed | 0.0022 | 29.1803 | 114.1306 |




---

## Problem 1.5 Discussion

1. **Answer:**

   Bias is low across all scenarios for both $\hat{\beta}$ and $\hat{\beta}_p$. The bias of both estimates decrease as $n$ increases. Bootstrap p CIs were more computationally demanding than the Wald CIs and performed comparably, or, particularly in the case of $n = 10$, worse than the other two methods. Bootstrap-$t$ CIs require considerable computational investment but performed best overall in the normal error scenarios and in the heavy-tailed error but small sample size ($n = 10$) scenarios. However, this had a trade-off in that it generally had somewhat wider CIs.

2. **Answers:**

- The Wald method is the fastest by far, followed by the nonparametric Bootstrap p method. The Bootstrap $t$ method is very slow across all scenarios. Computation time was consistent across all scenarios with Wald method taking a fraction of a second, Bootstrap p taking around 30 seconds, and

Bootstrap-$t$ taking around 105 seconds for $n = 10, 50$, with a slight bump to around the 115 seconds for $n = 500$.

- In terms of coverage, Bootstrap p performed noticably worse than the other two methods for the $n = 10$ and $n = 50$ scenarios but comparably for the $n = 500$ scenario.

  Overall I would say Bootstrap-$t$ had the best coverage of the methods, particularly for small sample sizes. However, the confidence intervals were much wider than that of other methods

  All methods struggled somewhat with scenario 2, which had a moderate sample size ($n = 50$) but no true effect ($\beta_t x = 0$)–none quite hit the 95% level.

- At very small sample sizes ($n = 10$), Bootstrap p has the worst coverage, which are also worse than the respective normal error scenarios. Wald and Bootstrap-$t$ CIs have comparible coverage at small sample size ($n = 10$). All methods seem to do a bit better for $\beta_{tx} = 2$ compared to the smaller values. Overall, in the heavy-tailed error scenarios the Wald CI have the best coverage.

  The plots of the actual confidence intervals show some very large outliers in the heavy-tailed error scenarios, particularly for $n = 10$ and worse for the Bootstrap-$t$ CIs.

The Bootstrap methods do not seem appropriate for very small sample sizes; the Bootstrap p CI has poor coverage while the Bootstrap-$t$ CI has better coverage but is much wider than other methods. Both would be computationally intensive for little benefit in these scenarios.