simulation_rst

Rui Hu

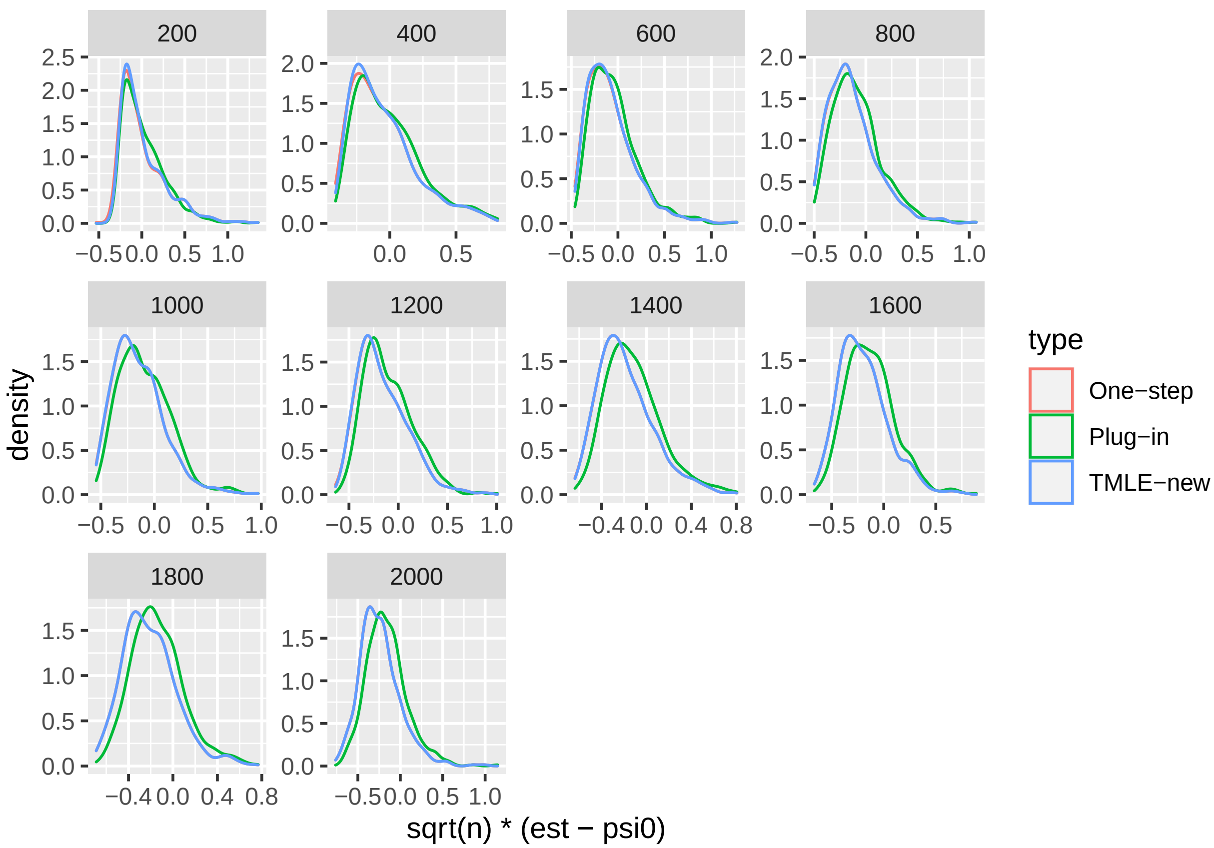
8/2/2021

Generating Y depending on A

Using SL.gam to estimate $\hat{\pi}$ and $\hat{\mu}$

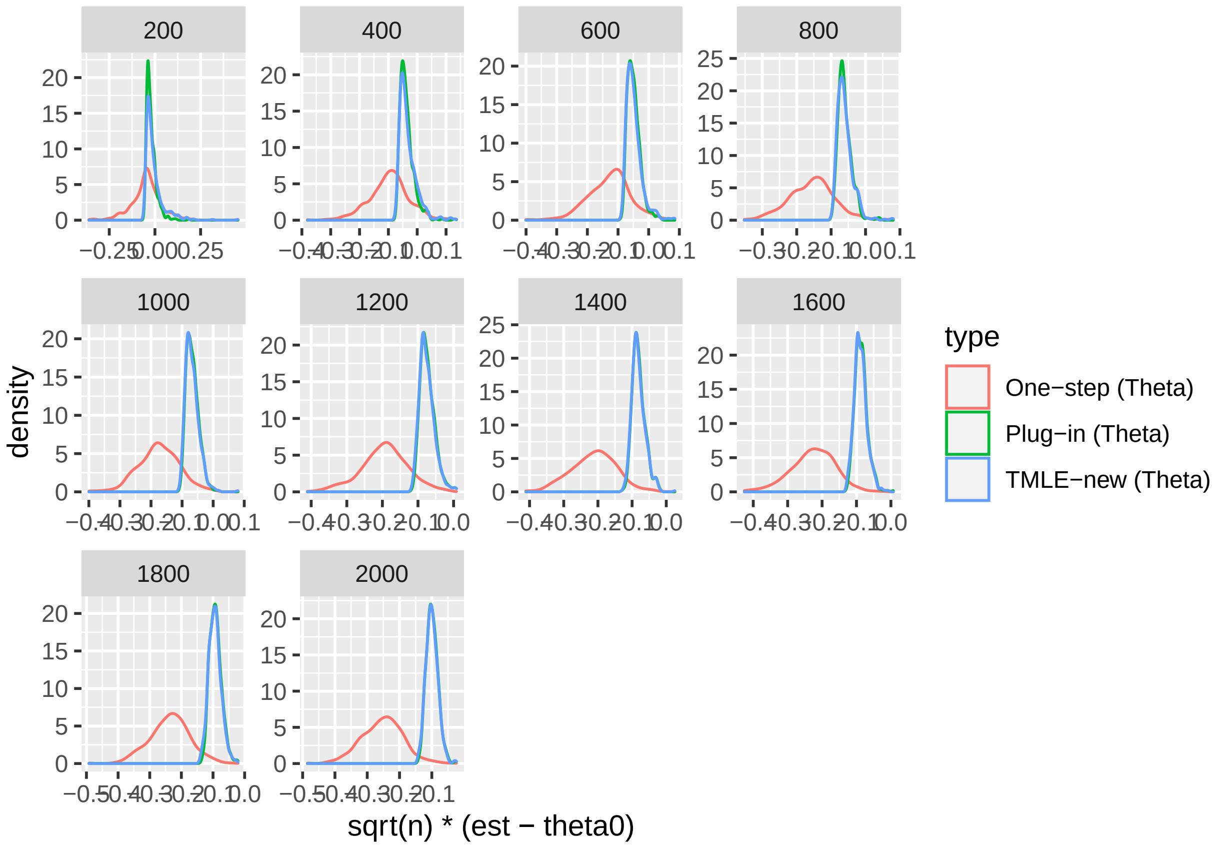
When n is large, `one-step` estimator of $\psi(P)$ tend to be non-negative.

```
# ests.sim.1 <-  est.psi.sim(200*c(1:10), 1:500,
#                            func_1 = "SL.gam", func_2 = "SL.gam", null.sims=FALSE)
load("ests.sim.1.RData")
est.psi.plot(ests.sim.1, plot.type='density')
```



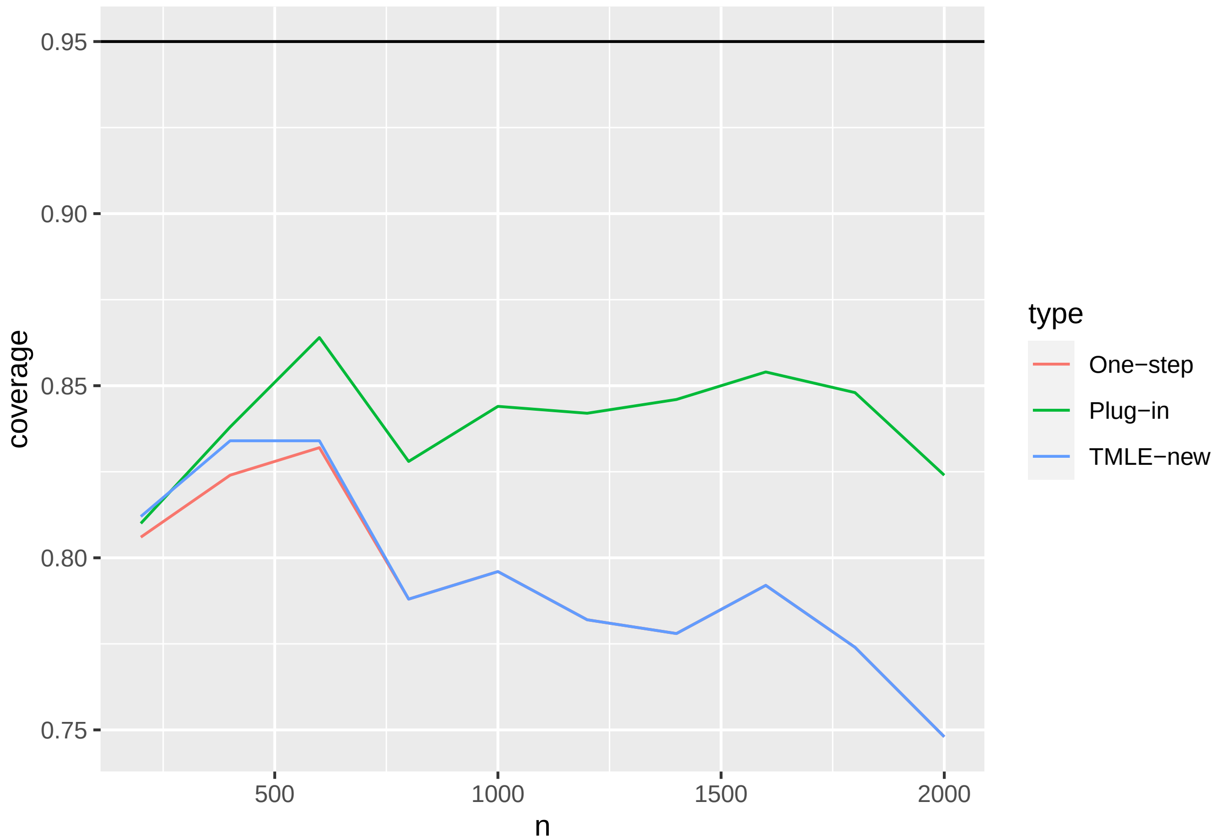
`one-step` estimator of $\theta(P)$ tend to be negative.

```
est.theta.plot(ests.sim.1, plot.type='density')
```

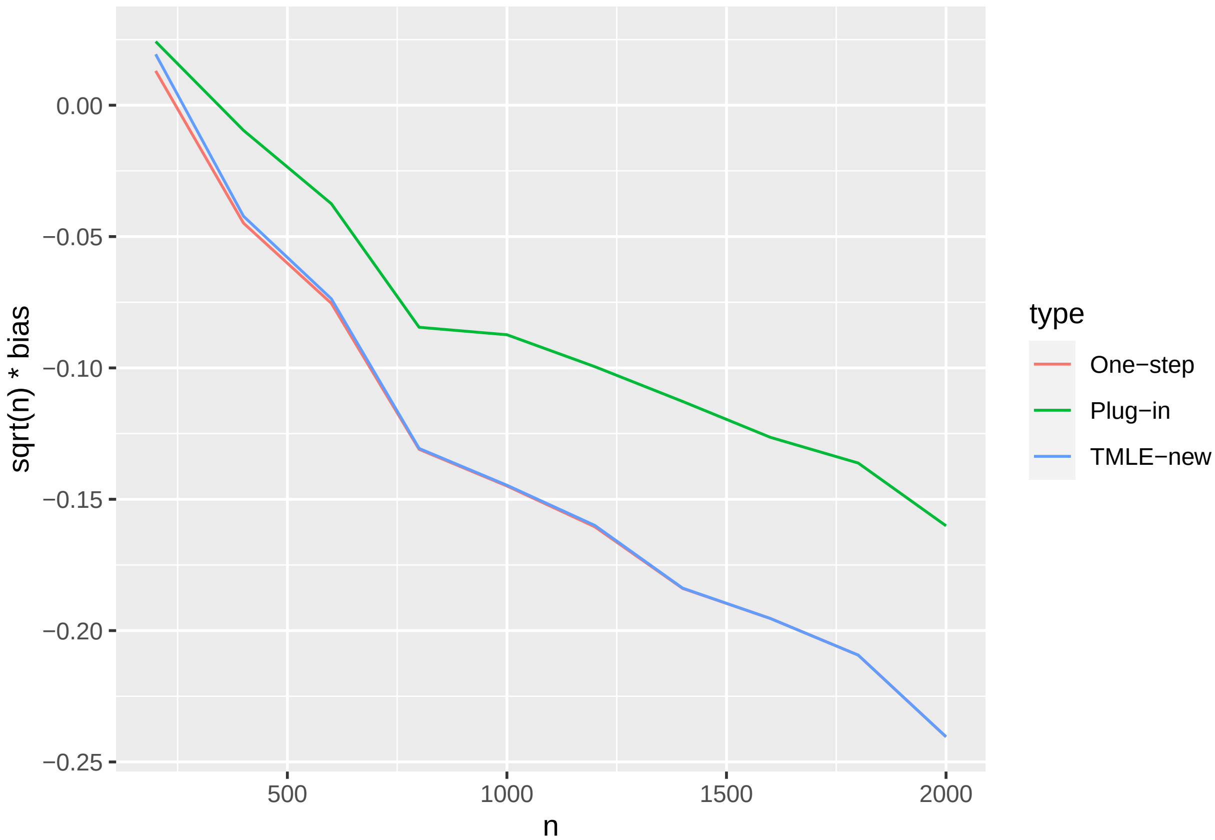


When n is increasing, the coverage is decreasing and the bias is increasing.

```
summaries.psi <- est.psi.summary(ests.sim.1)
ggplot(summaries.psi) +
  geom_line(aes(n, coverage, color=type)) +
  geom_hline(yintercept=.95)
```

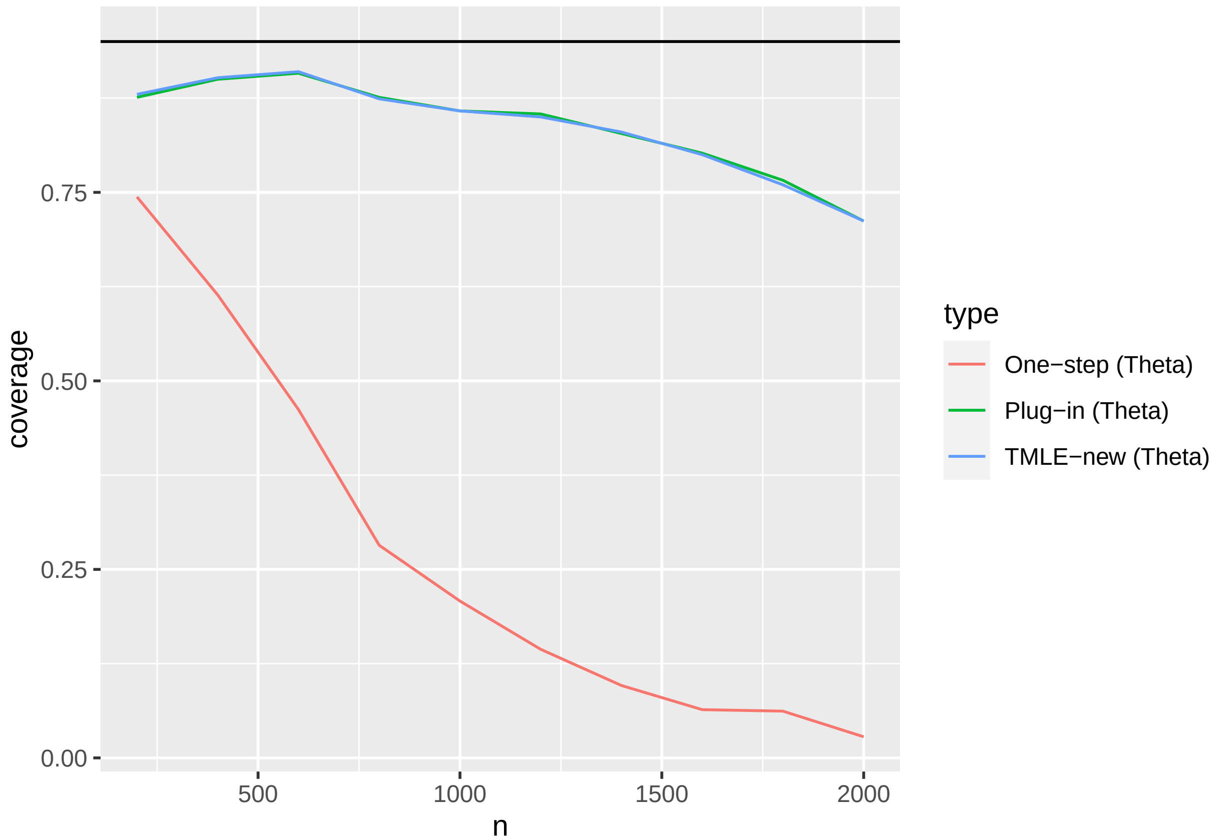



```r
ggplot(summaries.psi) +
  geom_line(aes(n, sqrt(n) * bias, color=type))
```



I am looking for possible explanations.
* [x] The MSE of gam model is the lowest among the three models. Thus the reason that leads to low coverage should not be the selection of model.

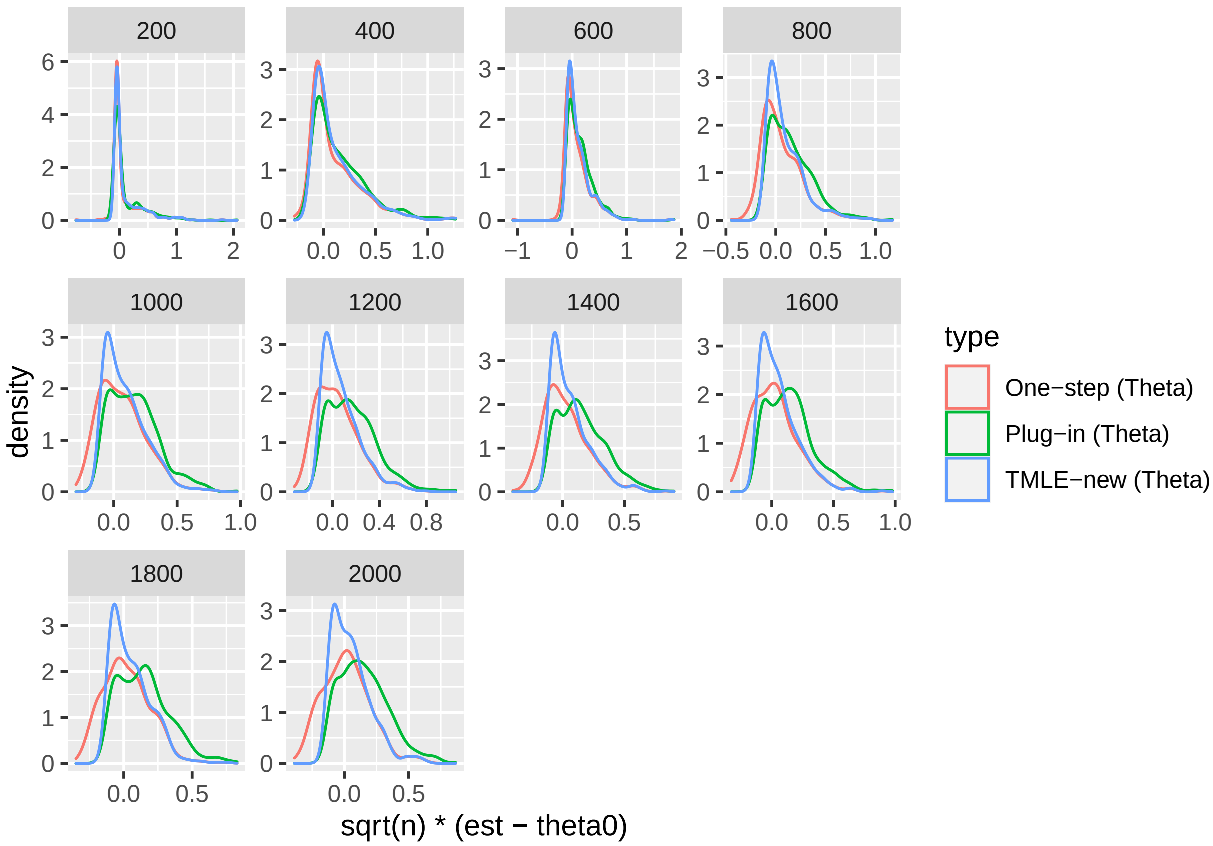
```
summaries.theta <- est.theta.summary(ests.sim.1)
ggplot(summaries.theta) +
  geom_line(aes(n, coverage, color=type)) +
  geom_hline(yintercept=.95)
```



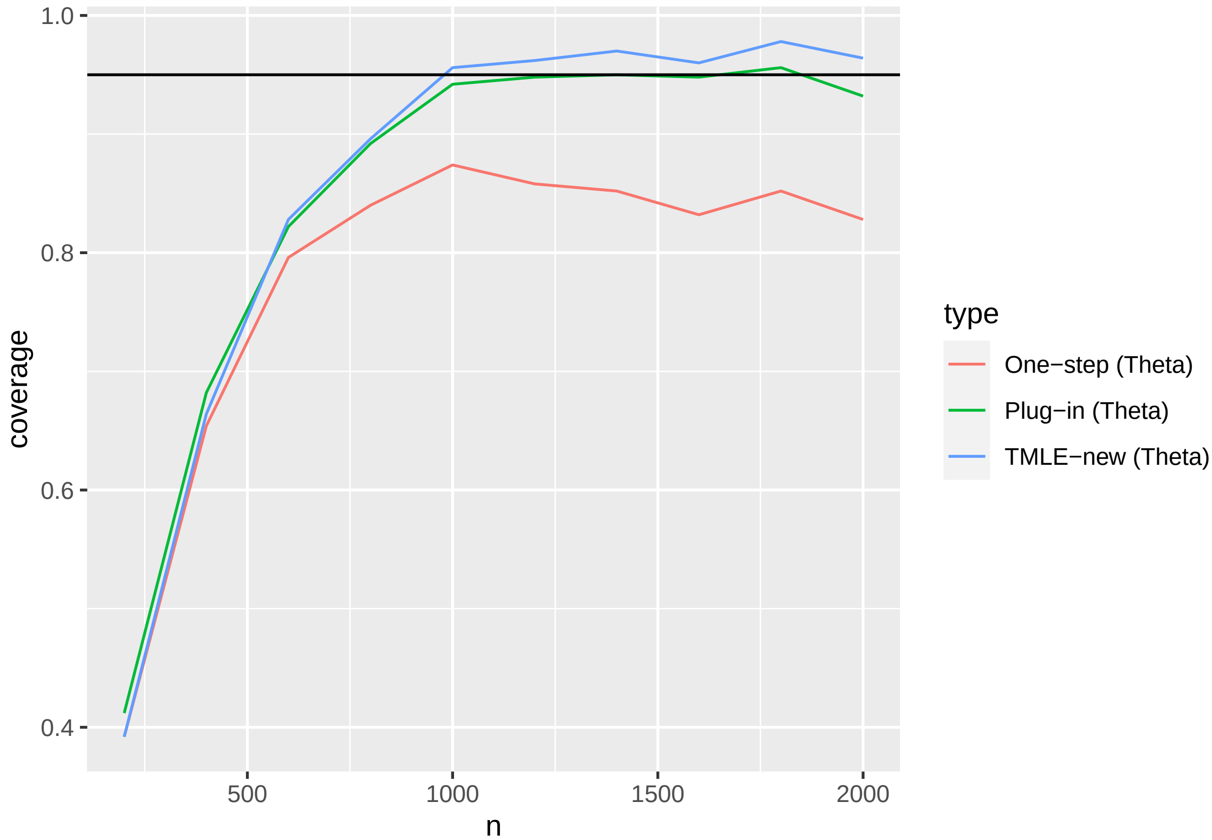
Using SL.earth to estimate $\hat{\pi}$ and $\hat{\mu}$

Since the model will do the variable selection, A might not be included to the estimated model and thus $\hat{\tau} = 0$ and $\hat{\theta}(P) = 0$.

```
load("ests.sim.3.RData")
est.theta.plot(ests.sim.3, plot.type='density')
```

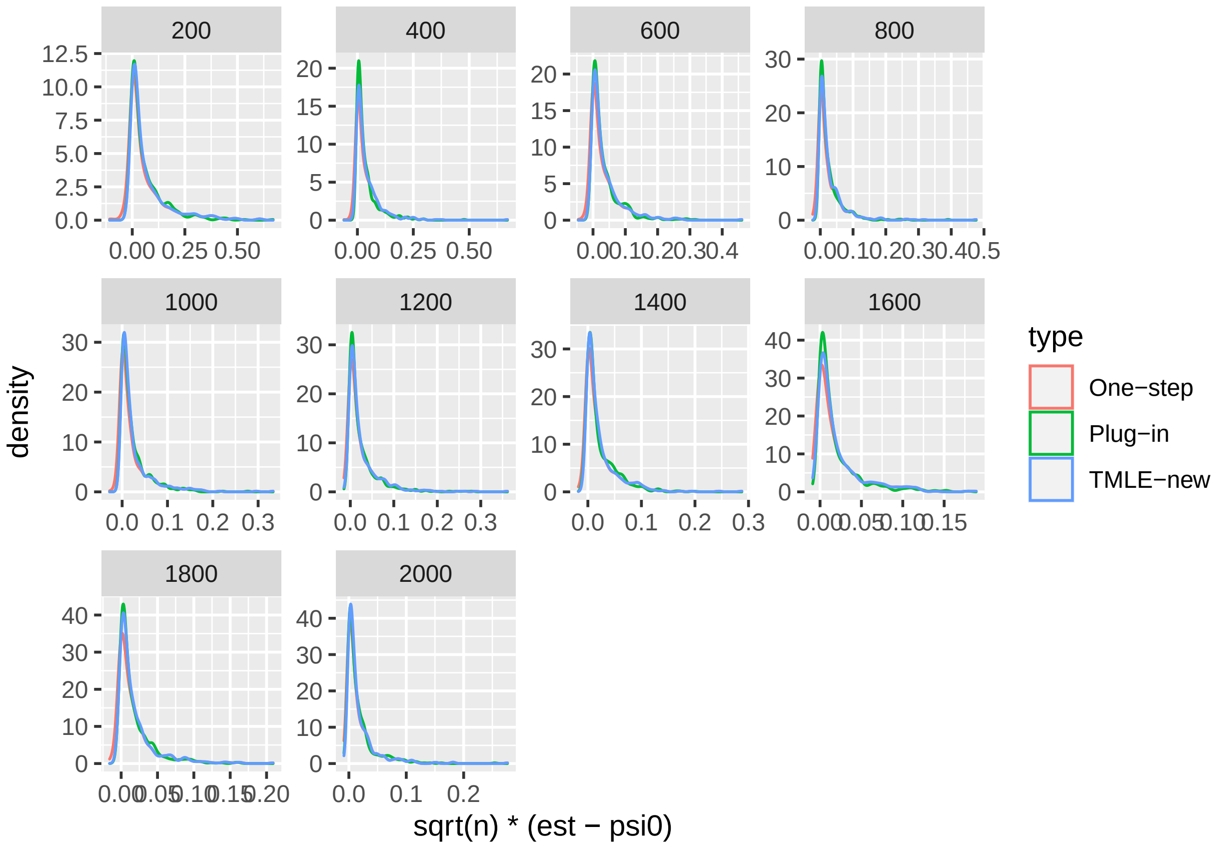



```
summaries.theta.3 <- est.theta.summary(ests.sim.3)
ggplot(summaries.theta.3) +
  geom_line(aes(n, coverage, color=type)) +
  geom_hline(yintercept=.95)
```

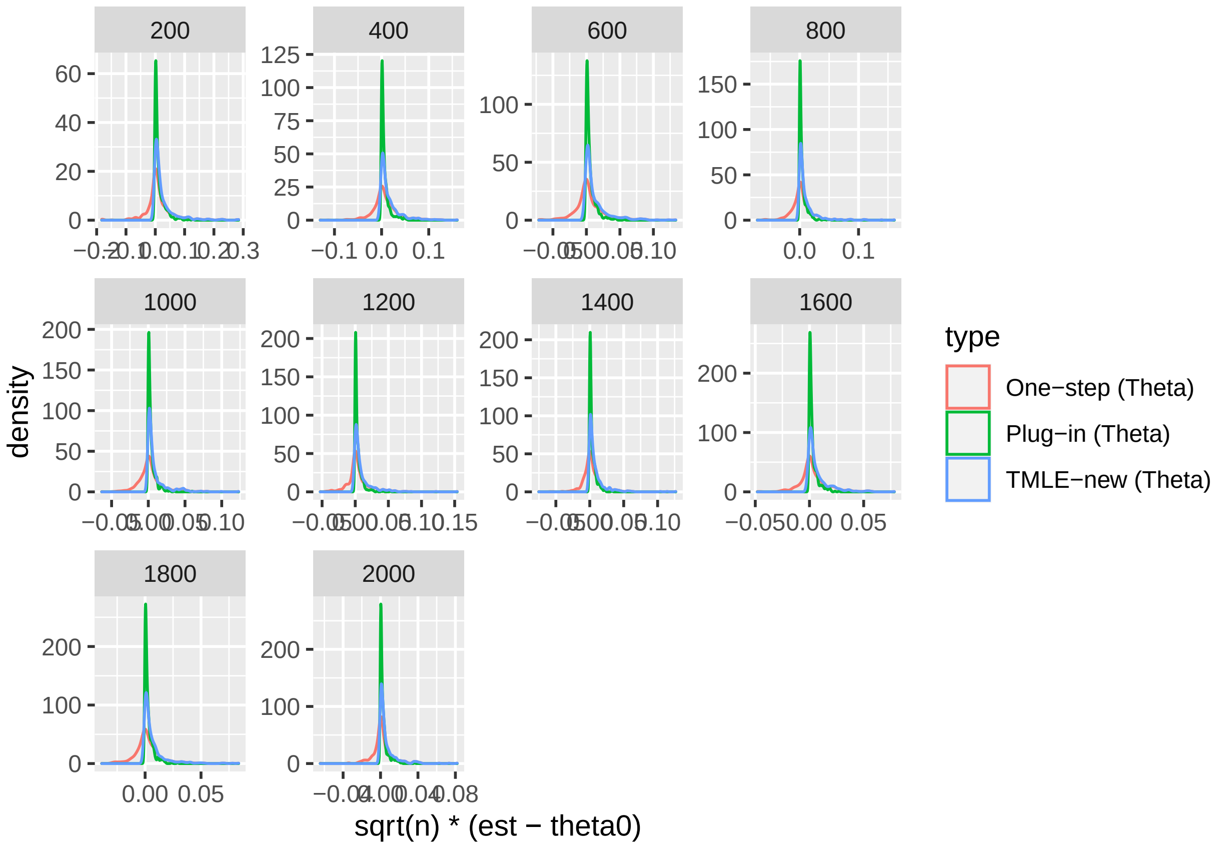


Generating Y not depending on A

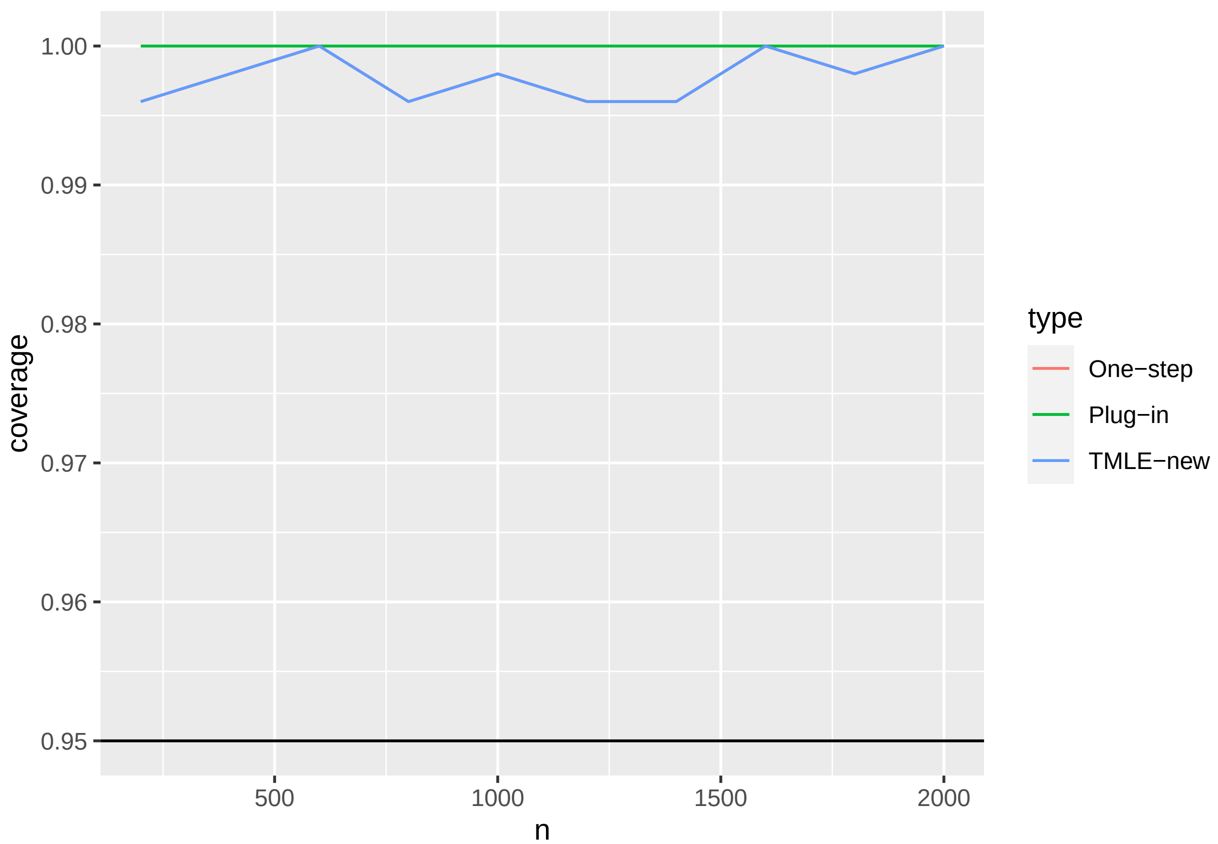
```
# ests.sim.1.null <-  est.psi.sim(200*c(1:10), 1:500,
#                              func_1 = "SL.gam", func_2 = "SL.gam", null.sims=TRUE)
load("ests.sim.1.null.RData")
est.psi.plot(ests.sim.1.null, plot.type='density')
```



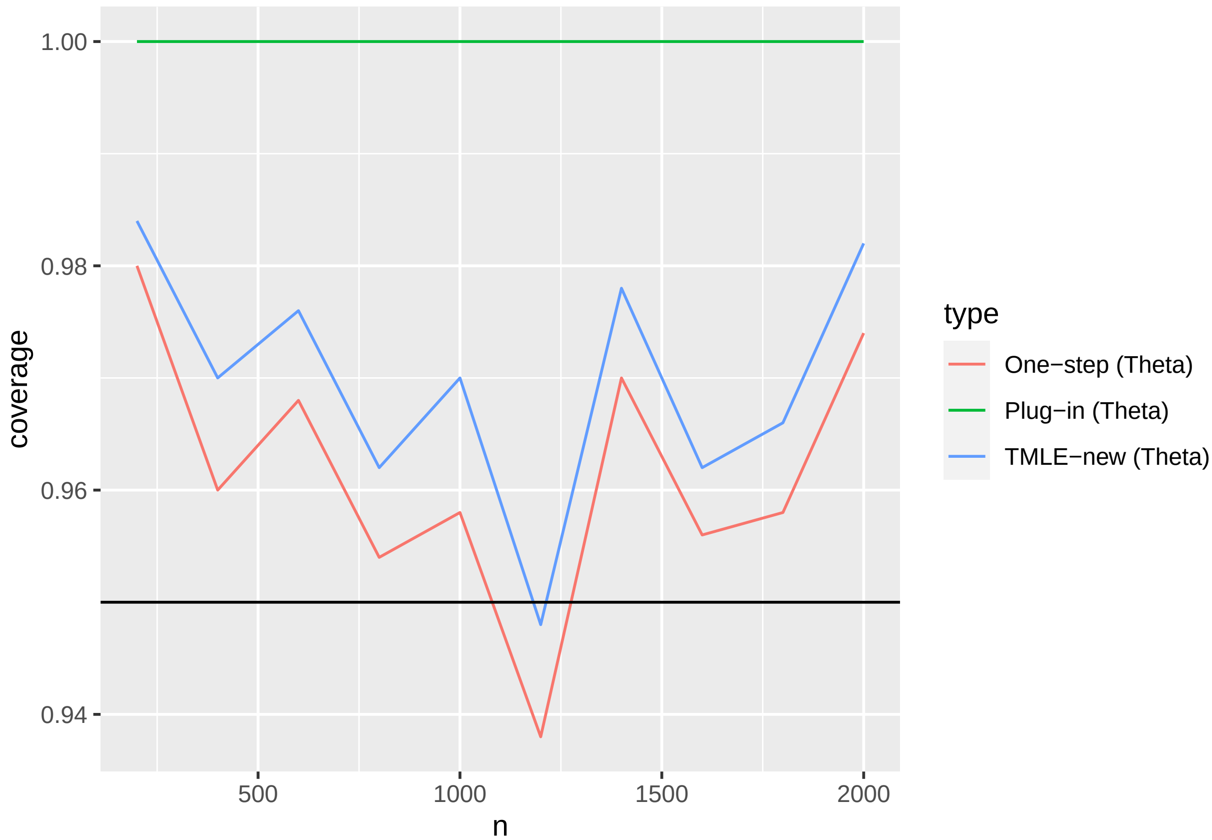

```r
est.theta.plot(ests.sim.1.null, plot.type='density')
```




```
summaries.psi.null <- est.psi.summary(ests.sim.1.null)
ggplot(summaries.psi.null) +
  geom_line(aes(n, coverage, color=type)) +
  geom_hline(yintercept=.95)
```




```
summaries.theta.null <- est.theta.summary(ests.sim.1.null)
ggplot(summaries.theta.null) +
  geom_line(aes(n, coverage, color=type)) +
  geom_hline(yintercept=.95)
```


1.00
0.98
0.96
0.94
coverage
500
1000
1500
2000
n
type
One−step (Theta)
Plug−in (Theta)
TMLE−new (Theta)