

Update

Matt Kosko

2024-08-13

METHOD 1 - Full Sample + BLB Bootstrapped SE (Wald CI Formed at End)

```
base_nm <- 'blb_standard_error_v1'
temp_dir <- file.path(base_dir, dat_path, paste0(base_nm, '_tmp'))
cblb <- as.data.table(readRDS(file.path(temp_dir, 'coverage.rds')))
cblb[, `:=`(lower_ci = tau_hat - qnorm(0.975)*boot_sd,
            upper_ci = tau_hat + qnorm(0.975)*boot_sd)][
  , .(coverage = mean(lower_ci <= te & upper_ci >= te)), by = c('n', 'subsets', 'prop_form
```

##	n	subsets	prop_form	out_form	coverage
## 1:	50000	10	correct	correct	0.951
## 2:	50000	50	correct	correct	0.951
## 3:	50000	1	correct	correct	0.948
## 4:	50000	10	wrong	correct	0.953
## 5:	50000	50	wrong	correct	0.955
## 6:	50000	1	wrong	correct	0.953
## 7:	50000	10	correct	wrong	0.954
## 8:	50000	50	correct	wrong	0.955
## 9:	50000	1	correct	wrong	0.956
## 10:	50000	10	wrong	wrong	0.000
## 11:	50000	50	wrong	wrong	0.000
## 12:	50000	1	wrong	wrong	0.000

METHOD 2 - Bias Correct BLB Replicates in Each Set, Take the Mean for an Estimate of the ATE, Wald Interval at the End

```
base_nm <- 'blb_standard_error_v2'
temp_dir <- file.path(base_dir, dat_path, paste0(base_nm, '_tmp'))
cblb <- as.data.table(readRDS(file.path(temp_dir, 'coverage.rds')))
cblb[, `:=`(lower_ci = tau_hat - qnorm(0.975)*boot_sd,
            upper_ci = tau_hat + qnorm(0.975)*boot_sd)][
  , .(coverage = mean(lower_ci <= te & upper_ci >= te)), by = c('n', 'subsets', 'prop_form
```

##	n	subsets	prop_form	out_form	coverage
## 1:	50000	10	correct	correct	0.951
## 2:	50000	50	correct	correct	0.951
## 3:	50000	1	correct	correct	0.948
## 4:	50000	10	wrong	correct	0.953
## 5:	50000	50	wrong	correct	0.955

```
## 6: 50000      1      wrong correct 0.953
## 7: 50000     10     correct wrong 0.954
## 8: 50000     50     correct wrong 0.955
## 9: 50000      1     correct wrong 0.956
## 10: 50000    10      wrong wrong 0.000
## 11: 50000    50      wrong wrong 0.000
## 12: 50000     1      wrong wrong 0.000
```

METHOD 3 - Bias Correct BLB Replicates, Take Percentile CI In Each Subset

```
base_nm <- 'blb_standard_error_v3'
temp_dir <- file.path(base_dir, dat_path, paste0(base_nm, '_tmp'))
cblb <- as.data.table(readRDS(file.path(temp_dir, 'coverage.rds')))
cblb[, .(coverage = mean(lower_ci <= te & upper_ci >= te)), by = c('n', 'subsets', 'prop_form', 'out_f
```

```
##          n subsets prop_form out_form coverage
## 1: 50000      10     correct correct 0.956
## 2: 50000      50     correct correct 0.958
## 3: 50000       1     correct correct 0.956
## 4: 50000      10      wrong correct 0.958
## 5: 50000      50      wrong correct 0.961
## 6: 50000       1      wrong correct 0.959
## 7: 50000      10     correct wrong 0.965
## 8: 50000      50     correct wrong 0.961
## 9: 50000       1     correct wrong 0.963
## 10: 50000     10      wrong wrong 0.000
## 11: 50000     50      wrong wrong 0.000
## 12: 50000      1      wrong wrong 0.000
```

Convergence of Each Replicate

```
base_nm <- 'convergence'
temp_dir <- file.path(base_dir, dat_path, paste0(base_nm, '_tmp'))
cblb <- as.data.table(readRDS(file.path(temp_dir, 'coverage.rds')))
cblb <- cblb[gamma != 0.5]

setorder(cblb, rep, subset_num)
cblb[, `:=`(blb_lower = cumsum(blb_lower)/subset_num, blb_upper = cumsum(blb_upper)/subset_num),
      by = c('rep', 'n', 'gamma', 'subsets', 'bias_correct')]
cblb[, `:=`(error = abs(true_lower - blb_lower)/2 + abs(true_upper - blb_upper)/2)]
trajectory <- cblb[, .(mean_error = mean(error)), by = c('n', 'gamma', 'subsets', 'bias_correct', 'subs
```

```
reps <- unique(cblb$rep)
```

```
for(i in reps){
  tmp <- cblb[rep == i]
```

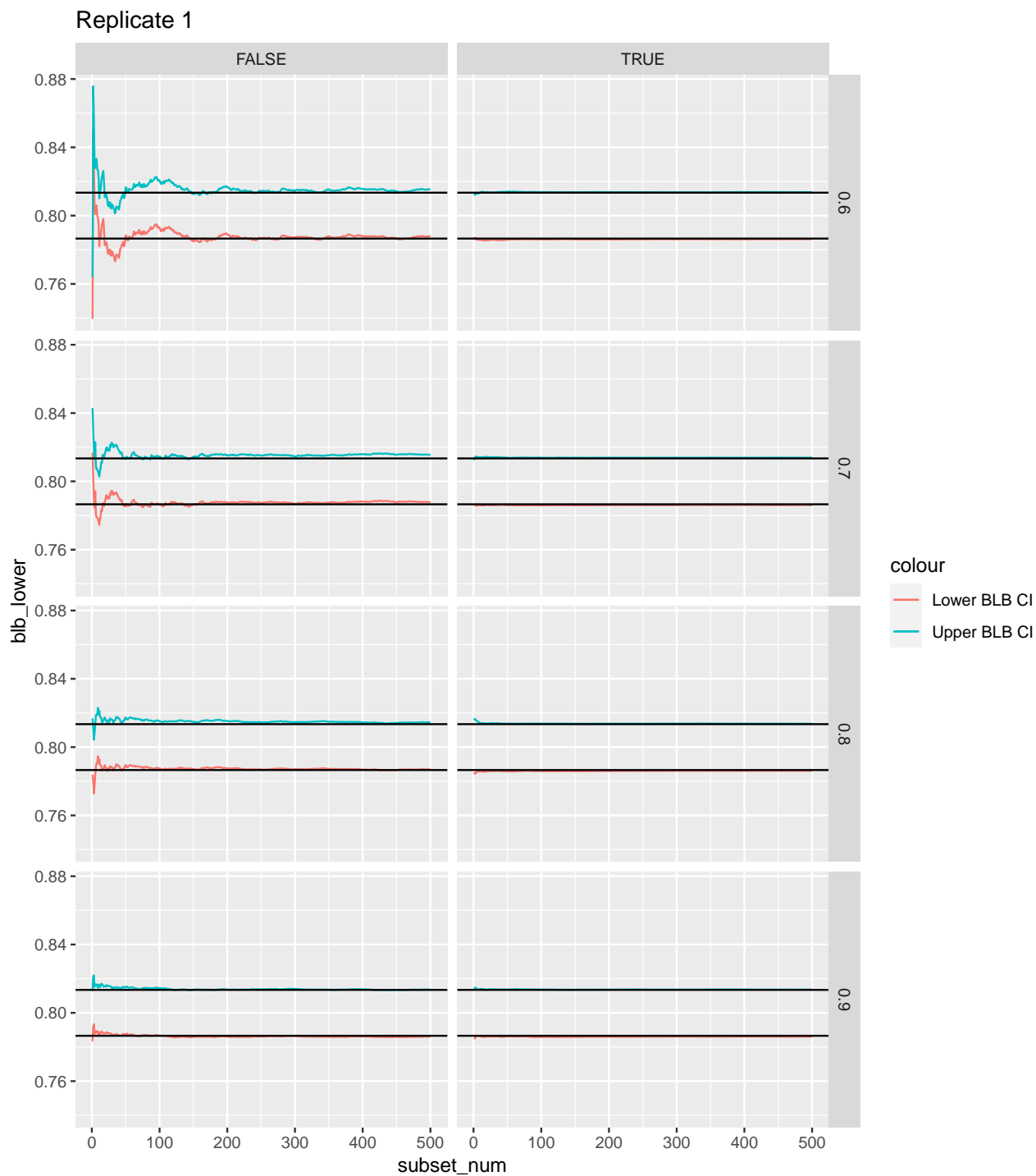
```
  p <- ggplot(tmp, aes(x = subset_num)) +
    geom_line(aes(y = blb_lower, color = 'Lower BLB CI')) +
    geom_line(aes(y = blb_upper, color = 'Upper BLB CI')) +
    geom_hline(aes(yintercept = true_lower)) +
```

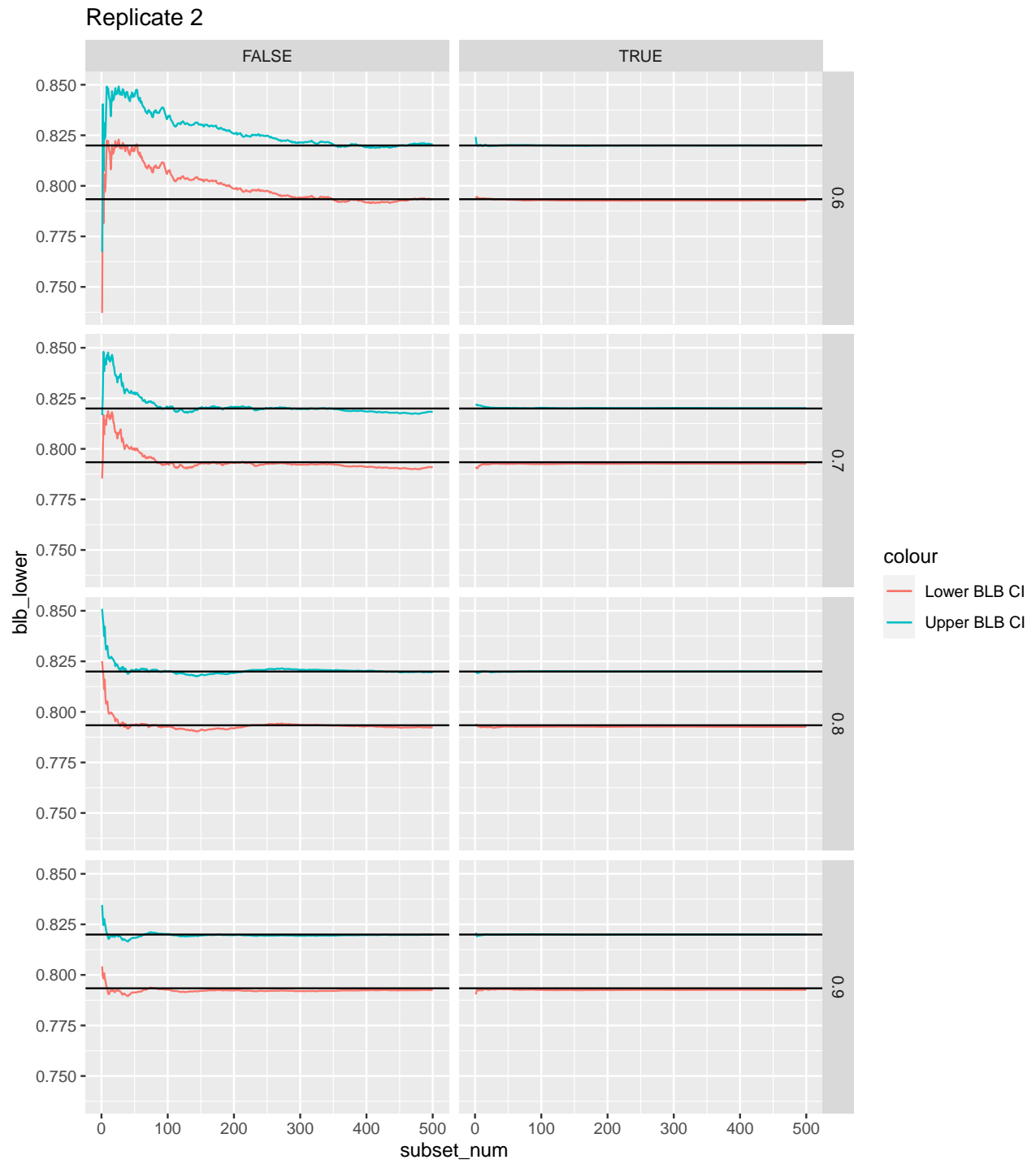
```

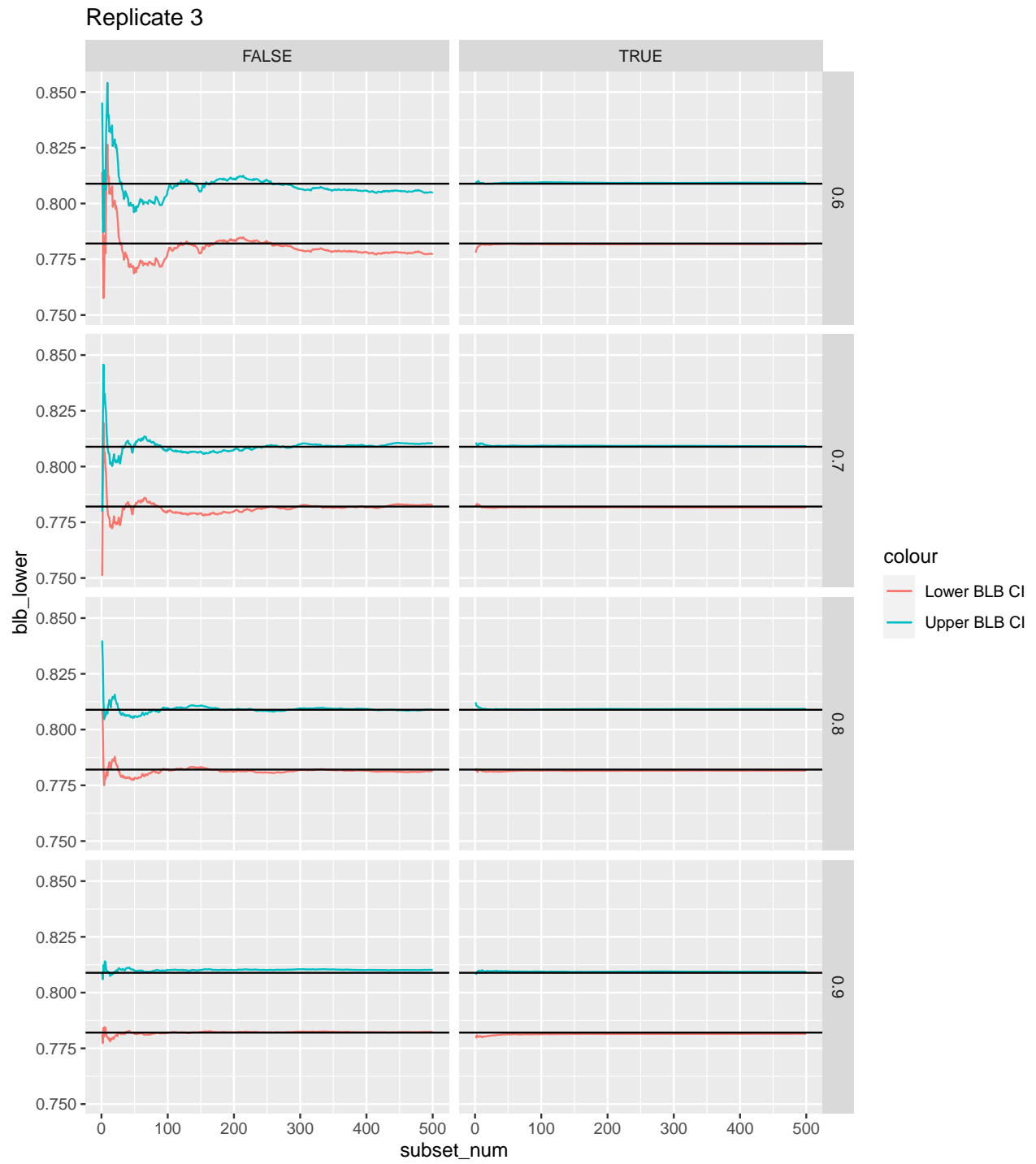
geom_hline(aes(yintercept = true_upper)) +
facet_grid(gamma ~ bias_correct) +
ggtitle(paste0('Replicate ', i))

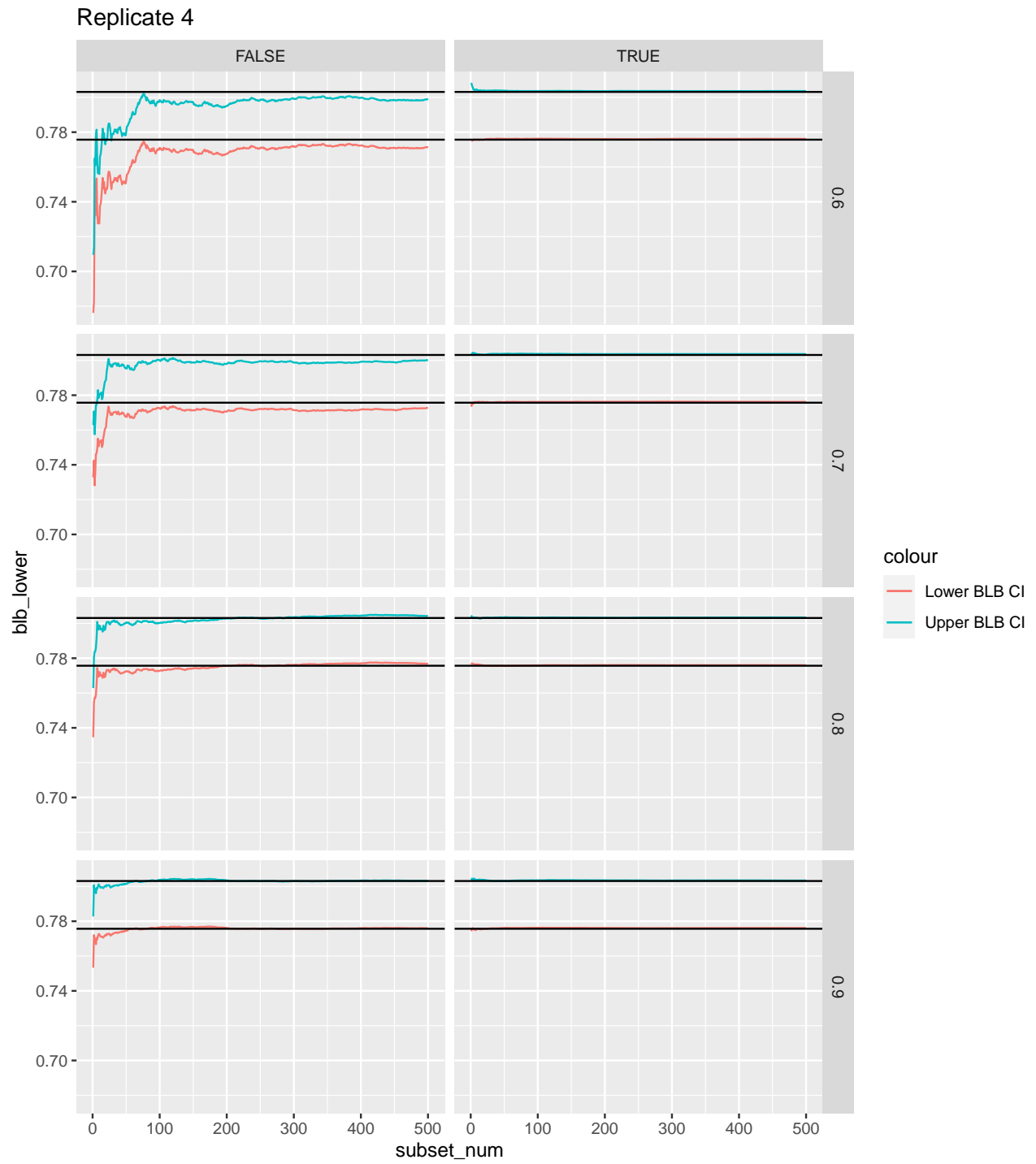
print(p)
}

```

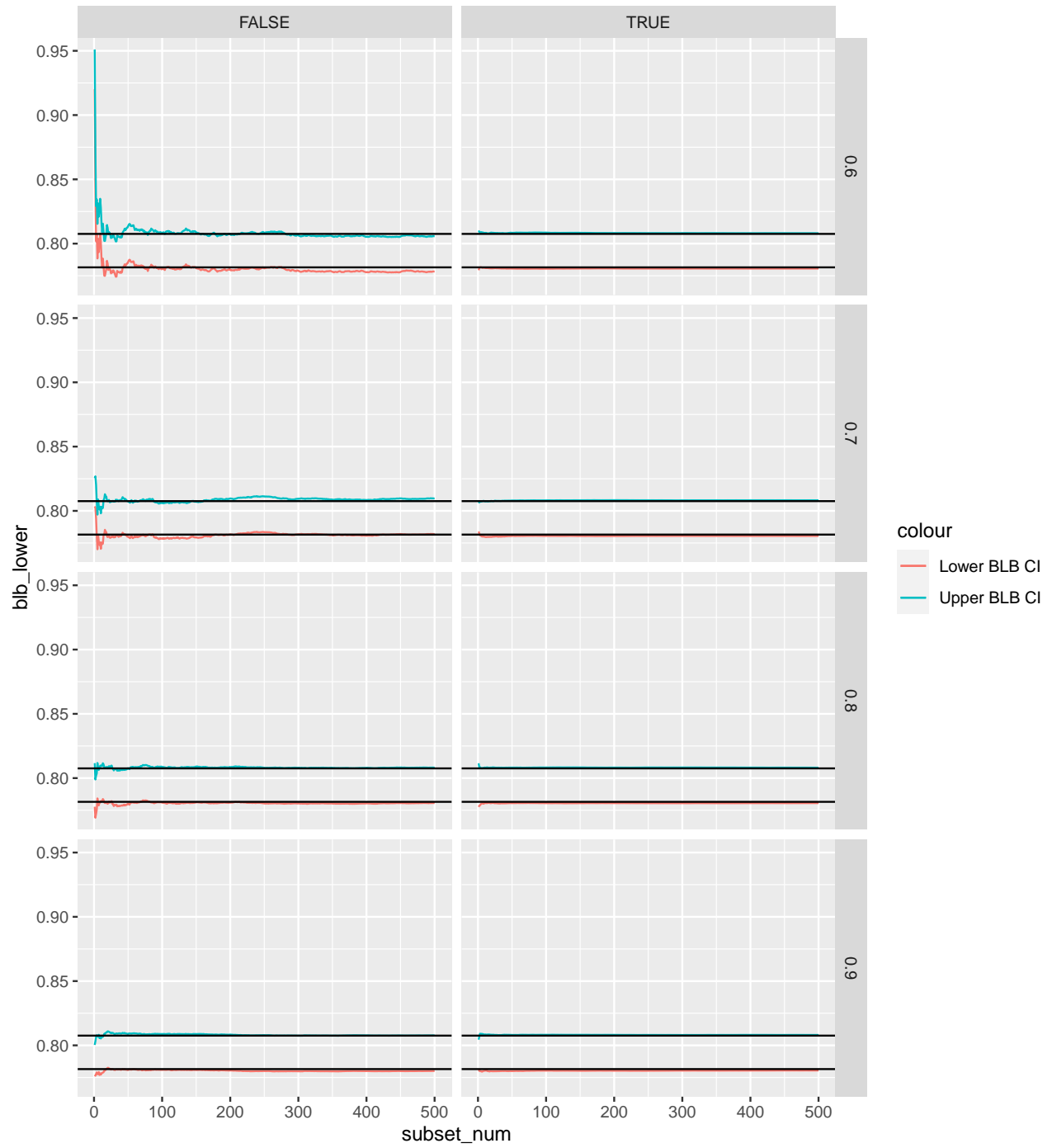


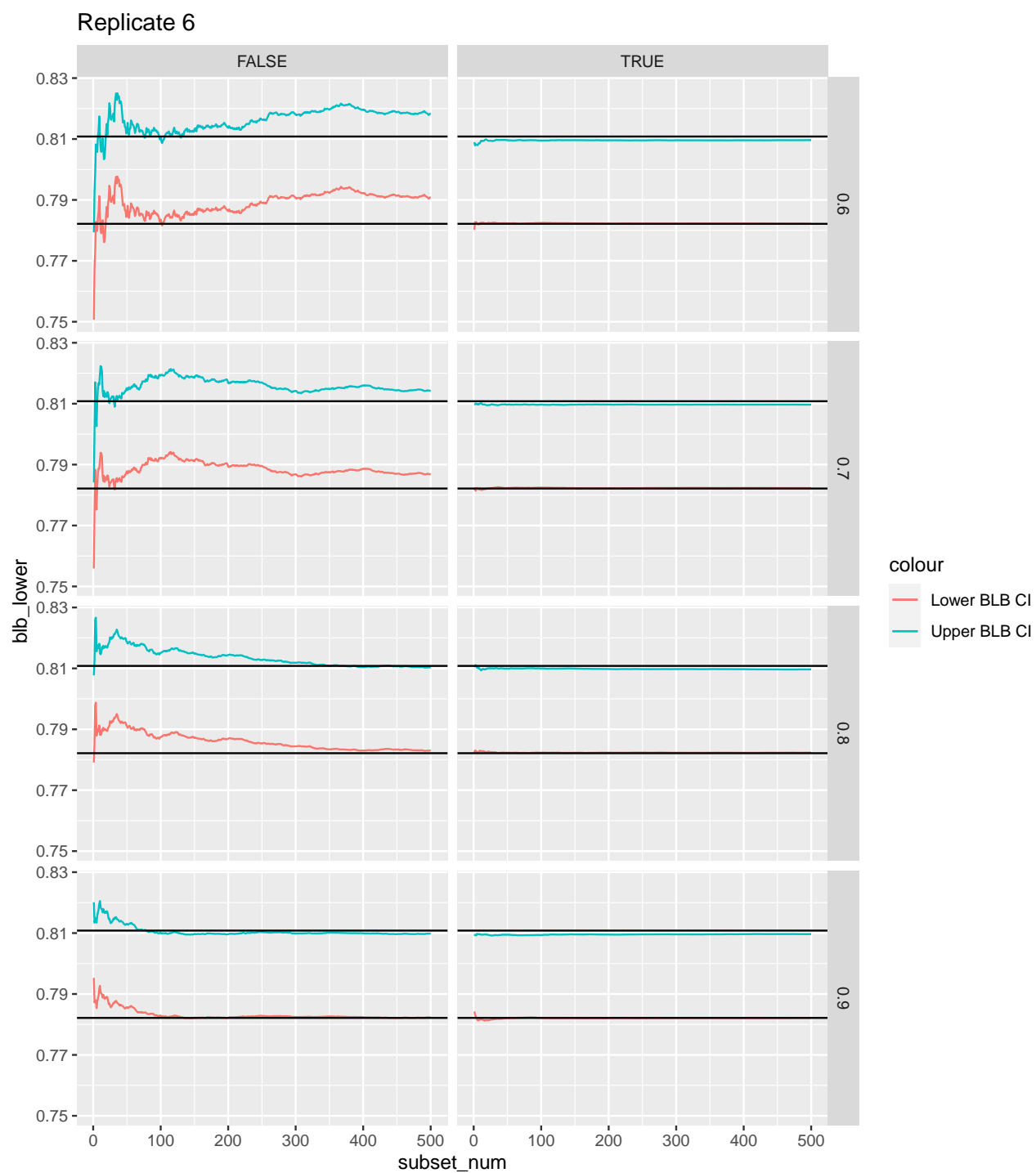


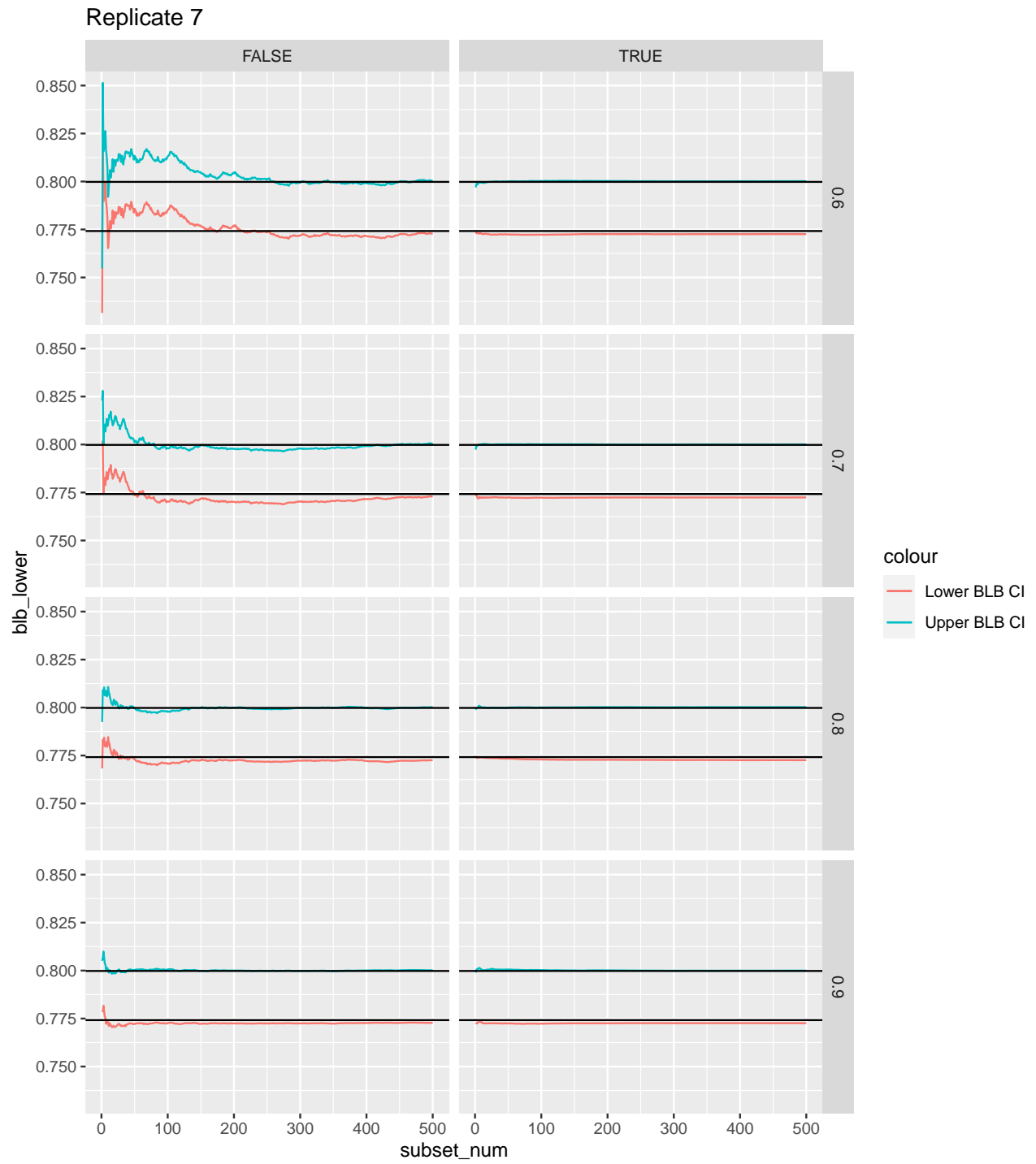




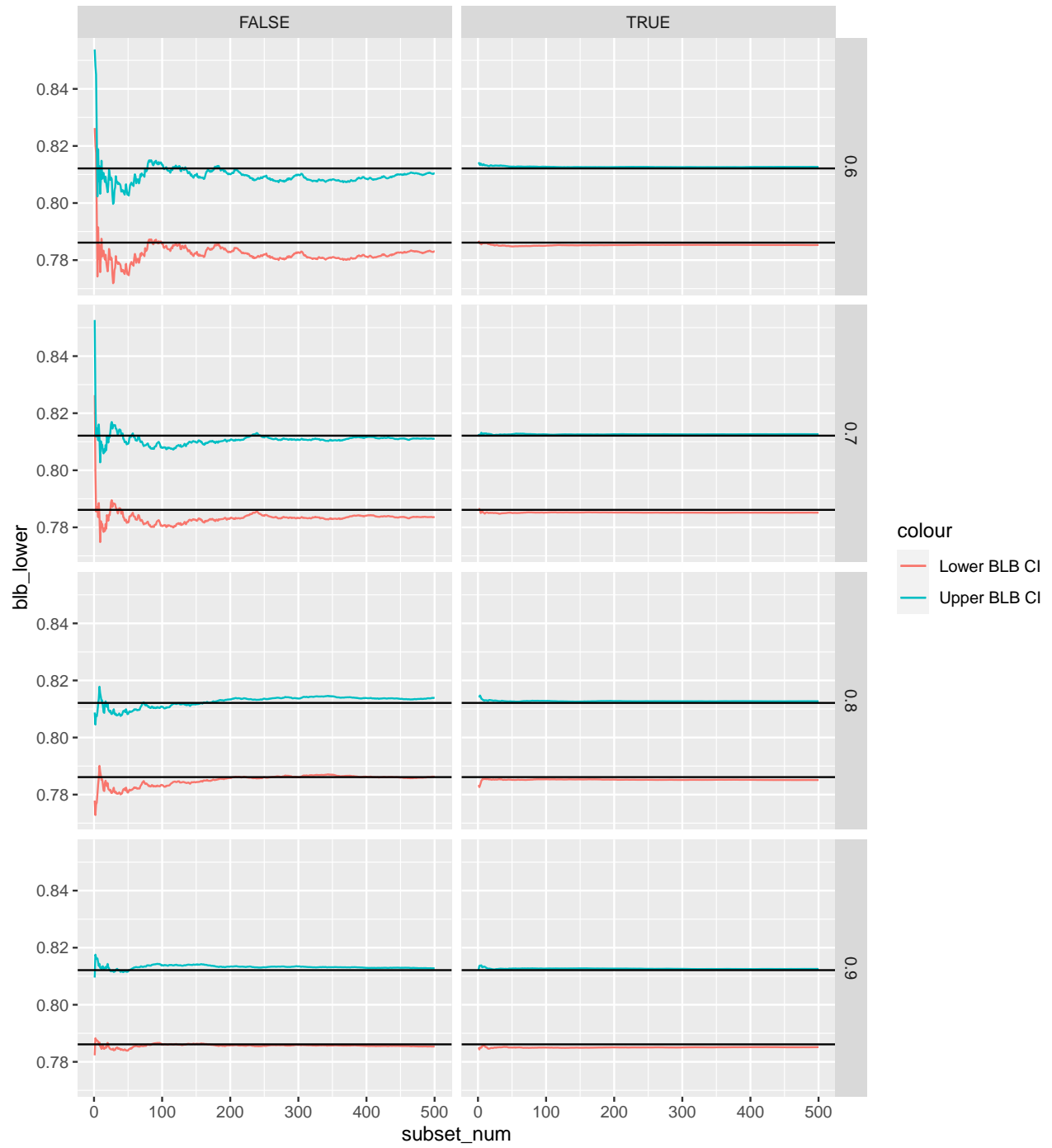
Replicate 5

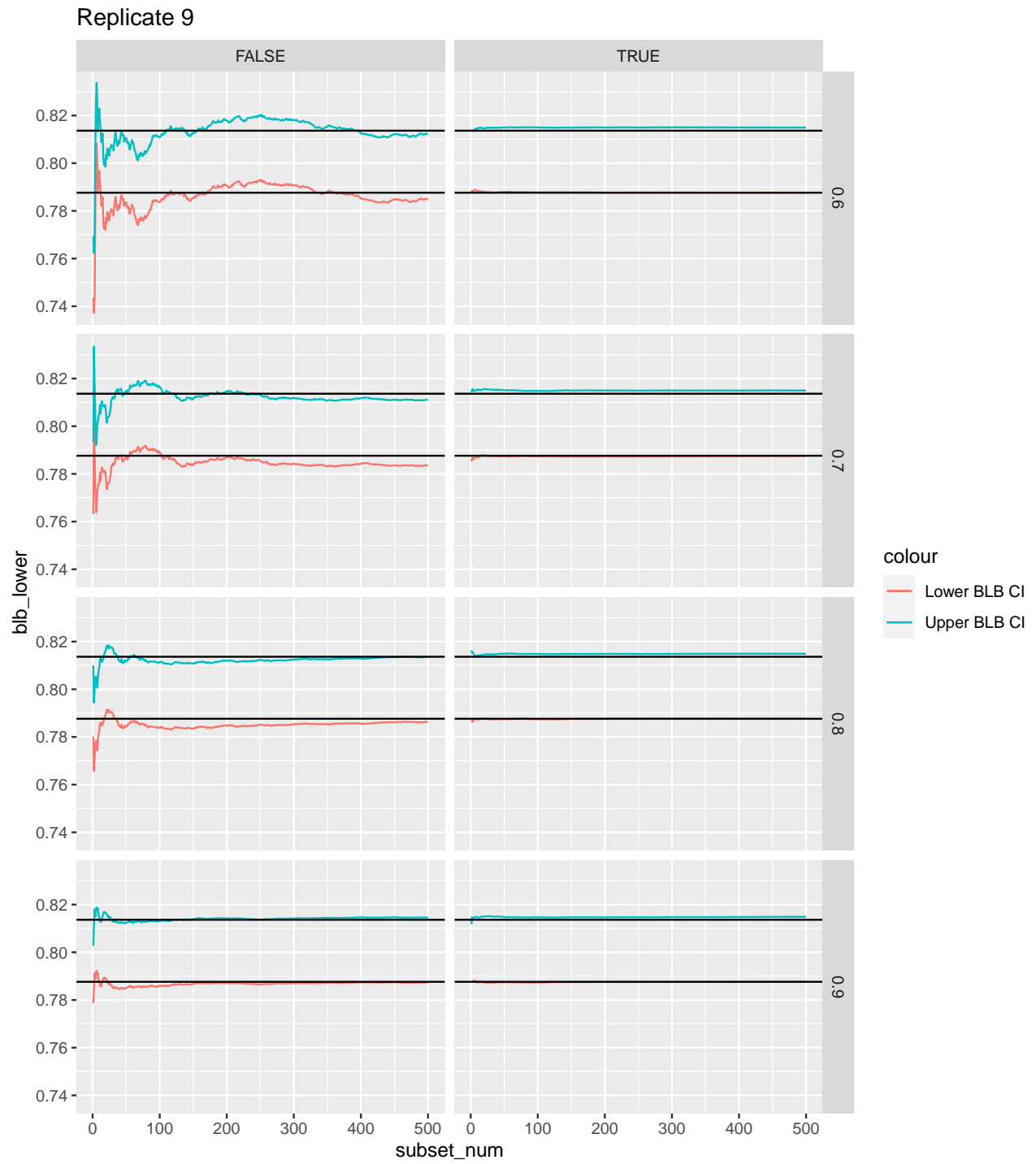


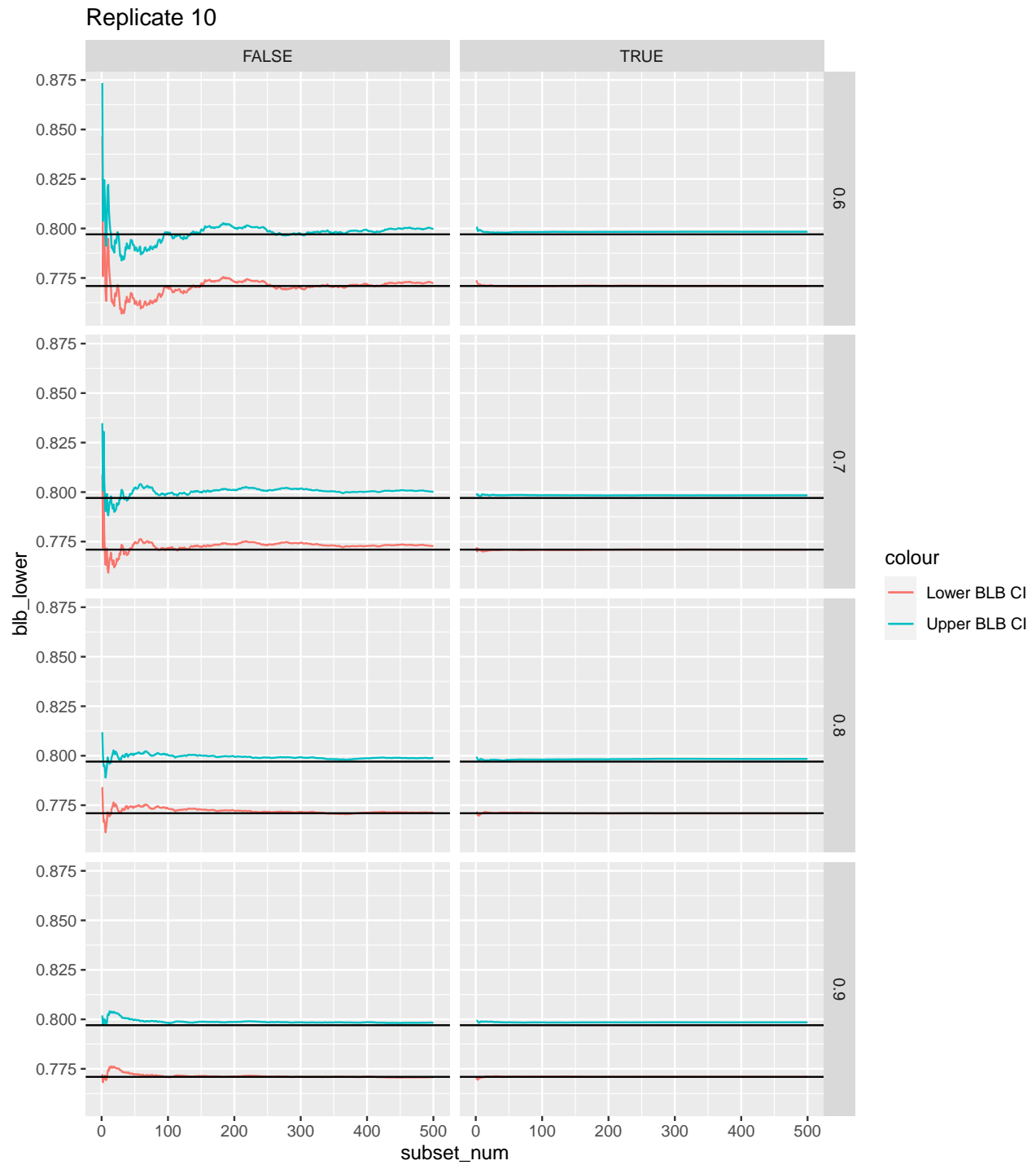




Replicate 8

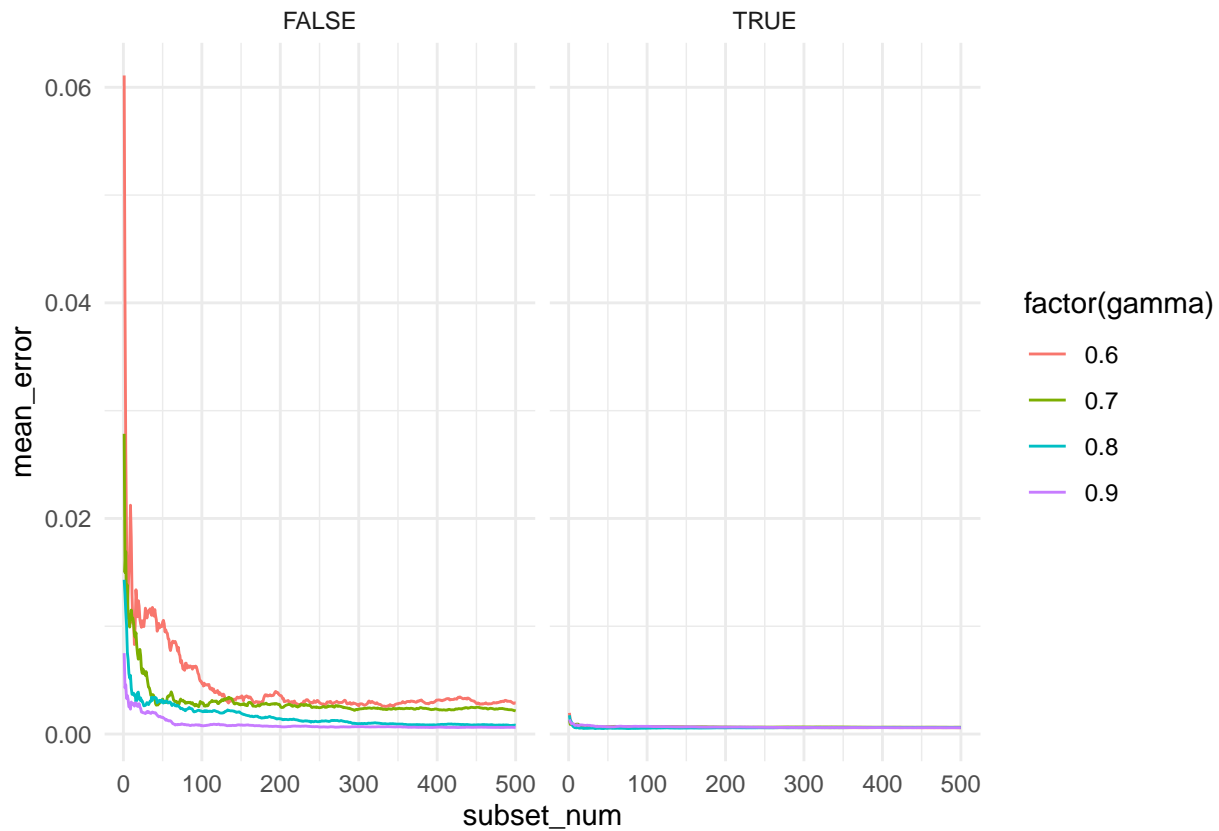






Error Convergence

```
ggplot(trajjectory, aes(x = subset_num, y = mean_error, color = factor(gamma))) +
  geom_line() +
  facet_grid(~ bias_correct) +
  theme_minimal()
```



Timing

```
base_nm <- 'blb_standard_error_ml_timing_full'
temp_dir <- file.path(base_dir, dat_path, paste0(base_nm, '_tmp'))
full <- as.data.table(readRDS(file.path(temp_dir, 'timing.rds')))
full[, `:=`(type = 'Regular bootstrap')]

base_nm <- 'blb_standard_error_ml_timing'
temp_dir <- file.path(base_dir, dat_path, paste0(base_nm, '_tmp'))
blb <- as.data.table(readRDS(file.path(temp_dir, 'timing.rds')))
blb[, `:=`(type = paste0('Cores: ', cores, '; Subsets = ', subsets, '; gamma = ', gamma))]
blb[, `:=`(cores = NULL, subsets = NULL, gamma = NULL)]

out <- rbindlist(list(blb, full))

ggplot(out, aes(x = type, y = time_elapsed)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90))
```

