Simulation study, misspecification

Tetiana Gorbach

Contents


```
load("../results/sim_study_misspecification_res2021_07_22_17_41_00.Rdata")
sim.results <- as.data.frame(output[[1]])
```

Transform to a long format:

```
sim.results.long <- sim.results %>%
  gather(method, ate.est, BD:`BD FD TD`, factor_key = TRUE) %>%
  mutate(
    `estimated ATE - true ATE` = ate.est - ate,
    method = factor(method, levels = c("FD", "FD TD", "BD", "BD TD", "TD", "BD FD TD")),
    misspecification = factor(misspecification,
      levels = c("Z", "C, A, Y", "C, Z", "A, Y"),
      labels = c("1. p(Z|A)", "2. All except p(Z|A)", "3. p(C), p(Z|A)", "4. p(A|C), E(Y|Z,C)")
    )
  )
```

Boxplots of estimates per model misspecification and estimation method:

```
cbbPalette <- c("#E69F00", "#D55E00", "#56B4E9", "#0072B2", "#009E73", "#000000", "#FF0000FF") # colorb
names(cbbPalette) <- levels(sim.results.long$method)

p <- sim.results.long %>%
  ggplot(aes(x = misspecification, y = `estimated ATE - true ATE`, fill = method)) +
  geom_boxplot() +
  scale_fill_manual(values = alpha(cbbPalette, .7)) +
  xlab("Model misspecification") +
  theme_bw() +
  theme(
    text = element_text(size = 6),
    axis.title = element_text(size = 8),
    strip.text = element_text(size = 8)
  ) +
  theme(
    legend.key.size = unit(2, "line"),
    legend.position = "bottom",
    legend.title = element_blank()
  ) +
  guides(fill = guide_legend(nrow = 1))

pdf("Figure_sim_study_misspecification.pdf", height = 4, width = 6)
p
dev.off()
```

```
## pdf
```

```
##   2
```

Simulation results

BIAS and MSE

```
sim.results.long <- sim.results.long %>%
  mutate(method = factor(method, levels = c("BD", "FD", "TD", "BD TD", "FD TD", "BD FD TD"))) # changin

s2 <- simsum(
  data = sim.results.long, estvarname = "ate.est", true = "ate", methodvar = "method", by = "misspecifi
  x = T
)
```

```
## 'ref' method was not specified, BD set as the reference
```

```
result.bias.se.mse <- summary(s2, digits = 3, ci.level = 0.95, stats = c("bias", "empse", "mse"))$summ
  mutate(estim = paste0(format(round(est, digits = 3), nsmall = 3), " (", format(round(mcse, digits = 3
  select(stat, estim, misspecification, method) %>%
  spread(method, estim)
result.bias.se.mse
```

```
##     stat     misspecification             BD             FD             TD
## 1   bias            1. p(Z|A)  0.000 (0.001)  0.000 (0.001)  0.000 (0.000)
## 2   bias 2. All except p(Z|A)  0.366 (0.001) -0.003 (0.002) -0.003 (0.001)
## 3   bias      3. p(C), p(Z|A)  0.000 (0.001)  0.000 (0.001)  0.000 (0.000)
## 4   bias  4. p(A|C), E(Y|Z,C)  0.000 (0.001) -0.001 (0.001) -0.003 (0.001)
## 5  empse            1. p(Z|A)  0.017 (0.000)  0.017 (0.000)  0.015 (0.000)
## 6  empse 2. All except p(Z|A)  0.018 (0.000)  0.055 (0.001)  0.046 (0.001)
## 7  empse      3. p(C), p(Z|A)  0.017 (0.000)  0.017 (0.000)  0.015 (0.000)
## 8  empse  4. p(A|C), E(Y|Z,C)  0.017 (0.000)  0.022 (0.000)  0.046 (0.001)
## 9    mse            1. p(Z|A)  0.000 (0.000)  0.000 (0.000)  0.000 (0.000)
## 10   mse 2. All except p(Z|A)  0.134 (0.000)  0.003 (0.000)  0.002 (0.000)
## 11   mse      3. p(C), p(Z|A)  0.000 (0.000)  0.000 (0.000)  0.000 (0.000)
## 12   mse  4. p(A|C), E(Y|Z,C)  0.000 (0.000)  0.000 (0.000)  0.002 (0.000)
##             BD TD          FD TD       BD FD TD
## 1   0.000 (0.000)  0.000 (0.000)  0.000 (0.000)
## 2  -0.048 (0.000) -0.003 (0.001) -0.048 (0.000)
## 3   0.000 (0.000) -0.091 (0.001) -0.091 (0.001)
## 4  -0.048 (0.000) -0.003 (0.001) -0.048 (0.000)
## 5   0.015 (0.000)  0.015 (0.000)  0.015 (0.000)
## 6   0.015 (0.000)  0.046 (0.001)  0.015 (0.000)
## 7   0.015 (0.000)  0.016 (0.000)  0.016 (0.000)
## 8   0.015 (0.000)  0.046 (0.001)  0.015 (0.000)
## 9   0.000 (0.000)  0.000 (0.000)  0.000 (0.000)
## 10  0.003 (0.000)  0.002 (0.000)  0.003 (0.000)
## 11  0.000 (0.000)  0.008 (0.000)  0.008 (0.000)
## 12  0.003 (0.000)  0.002 (0.000)  0.003 (0.000)
```

Scaled empirical variance vs bounds

```
result.sc.emp.var <- sim.results.long %>%
  group_by(misspecification, method) %>%
  summarise(sc.emp.var = var(sqrt(sample.size) * ate.est)) %>%
  ungroup() %>%
  mutate(mc.se.of.sc.emp.var = sqrt(2 * sc.emp.var^2 / number.of.replicates)) %>%
  mutate(estim = paste0(format(round(sc.emp.var, digits = 3), nsmall = 3), " (", format(round(mc.se.of.s
  select(misspecification, method, estim) %>%
  spread(method, estim) %>%
  add_column(stat = "ScEmpSE^2", .before = "misspecification") %>%
  add_row(tibble_row(
    stat = "Bound", misspecification = "",
    BD = as.character(round(output[[2]]["bound.BD"], digits = 3)),
    FD = as.character(round(output[[2]]["bound.FD"], digits = 3)),
    TD = as.character(round(output[[2]]["bound.TD"], digits = 3)),
    `BD TD` = as.character(round(output[[2]]["bound.BDTD"], digits = 3)),
    `FD TD` = as.character(round(output[[2]]["bound.FDTD"], digits = 3)),
    `BD FD TD` = as.character(round(output[[2]]["bound.BDFDTD"], digits = 3))
  ))
result.sc.emp.var
```

```
## # A tibble: 5 x 8
##   stat   misspecification  BD      FD      TD      `BD TD`  `FD TD`  `BD FD TD`
##   <chr>  <chr>             <chr>   <chr>   <chr>   <chr>    <chr>    <chr>
## 1 ScEmp~ "1. p(Z|A)"       " 14.5~ " 14.0~ " 11.8~ " 11.85~ " 11.54~ " 11.547 (~
## 2 ScEmp~ "2. All except p~ " 16.6~ "151.0~ "106.2~ " 11.00~ "106.25~ " 11.008 (~
## 3 ScEmp~ "3. p(C), p(Z|A)" " 14.5~ " 14.0~ " 11.8~ " 11.85~ " 12.81~ " 12.810 (~
## 4 ScEmp~ "4. p(A|C), E(Y|~ " 14.5~ " 23.2~ "106.2~ " 11.00~ "106.25~ " 11.008 (~
## 5 Bound  ""                "14.76~ "22.5"  "18.71" "11.864" "17.995" "11.15"
```

Printing for LaTeX:

```
print(xtable(
  rbind(result.bias.se.mse, result.sc.emp.var),
  align = rep("r", 9)
),
include.rownames = F
)
```

```
## % latex table generated in R 3.5.2 by xtable 1.8-4 package
## % Fri Jul 23 11:04:50 2021
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrrrrr}
##   \hline
## stat & misspecification & BD & FD & TD & BD TD & FD TD & BD FD TD \\
##   \hline
## bias & 1. p(Z$|$A) &  0.000 (0.001) &  0.000 (0.001) &  0.000 (0.000) &  0.000 (0.000) &  0.000 (0.0
##   bias & 2. All except p(Z$|$A) &  0.366 (0.001) & -0.003 (0.002) & -0.003 (0.001) & -0.048 (0.000) &
##   bias & 3. p(C), p(Z$|$A) &  0.000 (0.001) &  0.000 (0.001) &  0.000 (0.000) &  0.000 (0.000) & -0.0
##   bias & 4. p(A$|$C), E(Y$|$Z,C) &  0.000 (0.001) & -0.001 (0.001) & -0.003 (0.001) & -0.048 (0.000)
##   empse & 1. p(Z$|$A) &  0.017 (0.000) &  0.017 (0.000) &  0.015 (0.000) &  0.015 (0.000) &  0.015 (0
```

```
##   empse & 2. All except p(Z$|$A) &  0.018 (0.000) &  0.055 (0.001) &  0.046 (0.001) &  0.015 (0.000)
##   empse & 3. p(C), p(Z$|$A) &  0.017 (0.000) &  0.017 (0.000) &  0.015 (0.000) &  0.015 (0.000) &  0
##   empse & 4. p(A$|$C), E(Y$|$Z,C) &  0.017 (0.000) &  0.022 (0.000) &  0.046 (0.001) &  0.015 (0.000)
##   mse & 1. p(Z$|$A) &  0.000 (0.000) &  0.000 (0.000) &  0.000 (0.000) &  0.000 (0.000) &  0.000 (0.0
##   mse & 2. All except p(Z$|$A) &  0.134 (0.000) &  0.003 (0.000) &  0.002 (0.000) &  0.003 (0.000) &
##   mse & 3. p(C), p(Z$|$A) &  0.000 (0.000) &  0.000 (0.000) &  0.000 (0.000) &  0.000 (0.000) &  0.0
##   mse & 4. p(A$|$C), E(Y$|$Z,C) &  0.000 (0.000) &  0.000 (0.000) &  0.002 (0.000) &  0.003 (0.000) &
##   ScEmpSE\verb|^|2 & 1. p(Z$|$A) &  14.569 (0.652) &  14.076 (0.630) &  11.854 (0.530) &  11.854 (0.5
##   ScEmpSE\verb|^|2 & 2. All except p(Z$|$A) &  16.610 (0.743) & 151.088 (6.757) & 106.257 (4.752) &
##   ScEmpSE\verb|^|2 & 3. p(C), p(Z$|$A) &  14.569 (0.652) &  14.076 (0.630) &  11.854 (0.530) &  11.85
##   ScEmpSE\verb|^|2 & 4. p(A$|$C), E(Y$|$Z,C) &  14.560 (0.651) &  23.227 (1.039) & 106.257 (4.752) &
##   Bound &  & 14.765 & 22.5 & 18.71 & 11.864 & 17.995 & 11.15 \\
##    \hline
## \end{tabular}
## \end{table}
```