Simulation study, misspecification

Tetiana Gorbach

Contents

```
load("../results/sim_study_misspecification_res2023_08_09_19_32_22.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 ACE - true ACE` = 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)")
    )
  )
```

Simulation results

Boxplots of estimates per model misspecification and estimation method:

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

p <- sim.results.long %>%
  ggplot(aes(x = misspecification, y = `estimated ACE - true ACE`, 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_study2.pdf", height = 4, width = 6)
p
dev.off()
```

```
## pdf
##   2
```

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")))
# changing order of the levels for tables

s2 <- simsum(
  data = sim.results.long, estvarname = "ate.est", true = "ate",
  methodvar = "method", by = "misspecification",
  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 = 4), nsmall = 3), "(",
                        format(round(mcse, digits = 3), nsmall = 3), ")")) %>%
  select(stat, estim, misspecification, method) %>%
  spread(method, estim)
result.bias.se.mse
```

```
##       stat      misspecification              BD             FD             TD
## 1     bias               1. p(Z|A) -0.0006(0.001) -0.0004(0.001) -0.0006(0.000)
## 2     bias 2. All except p(Z|A)  0.3659(0.001) -0.0012(0.002) -0.0012(0.001)
## 3     bias       3. p(C), p(Z|A) -0.0006(0.001) -0.0004(0.001) -0.0006(0.000)
## 4     bias  4. p(A|C), E(Y|Z,C) -0.0006(0.001) -0.0008(0.001) -0.0012(0.001)
## 5    empse               1. p(Z|A)  0.0172(0.000)  0.0164(0.000)  0.0153(0.000)
## 6    empse 2. All except p(Z|A)  0.0180(0.000)  0.0558(0.001)  0.0472(0.001)
## 7    empse       3. p(C), p(Z|A)  0.0172(0.000)  0.0164(0.000)  0.0153(0.000)
## 8    empse  4. p(A|C), E(Y|Z,C)  0.0171(0.000)  0.0212(0.000)  0.0472(0.001)
## 9      mse               1. p(Z|A)  0.0003(0.000)  0.0003(0.000)  0.0002(0.000)
## 10     mse 2. All except p(Z|A)  0.1342(0.000)  0.0031(0.000)  0.0022(0.000)
## 11     mse       3. p(C), p(Z|A)  0.0003(0.000)  0.0003(0.000)  0.0002(0.000)
## 12     mse  4. p(A|C), E(Y|Z,C)  0.0003(0.000)  0.0005(0.000)  0.0022(0.000)
##            BD TD          FD TD       BD FD TD
## 1  -0.0006(0.000) -0.0005(0.000) -0.0005(0.000)
## 2  -0.0480(0.000) -0.0012(0.001) -0.0480(0.000)
## 3  -0.0006(0.000) -0.0919(0.000) -0.0919(0.000)
## 4  -0.0480(0.000) -0.0012(0.001) -0.0480(0.000)
## 5   0.0153(0.000)  0.0150(0.000)  0.0150(0.000)
```

```
## 6    0.0147(0.000)    0.0472(0.001)    0.0147(0.000)
## 7    0.0153(0.000)    0.0145(0.000)    0.0145(0.000)
## 8    0.0147(0.000)    0.0472(0.001)    0.0147(0.000)
## 9    0.0002(0.000)    0.0002(0.000)    0.0002(0.000)
## 10   0.0025(0.000)    0.0022(0.000)    0.0025(0.000)
## 11   0.0002(0.000)    0.0086(0.000)    0.0086(0.000)
## 12   0.0025(0.000)    0.0022(0.000)    0.0025(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 = 4), nsmall = 3), "(",
                        format(round(mc.se.of.sc.emp.var, digits = 3), nsmall = 3), ")")) %>%
  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))
  ))
```

```
## `summarise()` has grouped output by 'misspecification'. You can override using
## the `.groups` argument.
```

```
result.sc.emp.var
```

```
## # A tibble: 5 x 8
##   stat      misspecification       BD        FD     TD     `BD TD` `FD TD` BD FD~1
##   <chr>     <chr>                  <chr>     <chr>  <chr>  <chr>   <chr>   <chr>
## 1 ScEmpSE^2 "1. p(Z|A)"            " 14.800~ " 13~  " 11~  " 11.7~ " 11.2~ " 11.2~
## 2 ScEmpSE^2 "2. All except p(Z|A)" " 16.141~ "155~  "111~  " 10.8~ "111.4~ " 10.8~
## 3 ScEmpSE^2 "3. p(C), p(Z|A)"      " 14.800~ " 13~  " 11~  " 11.7~ " 10.5~ " 10.5~
## 4 ScEmpSE^2 "4. p(A|C), E(Y|Z,C)"  " 14.558~ " 22~  "111~  " 10.8~ "111.4~ " 10.8~
## 5 Bound     ""                     "14.765"  "22.~  "18.~  "11.86~ "17.99~ "11.15"
## # ... with abbreviated variable name 1: `BD FD TD`
```

Printing for LaTeX

```
print(xtable(
  rbind(result.bias.se.mse, result.sc.emp.var)%>%
        mutate(stat = recode(stat,  "bias" ="Bias", "empse" = "EmpSE", "mse" = "MSE", "ScEmpSE^2" = "ScE
  align = rep("r", 9)
),
include.rownames = F
)
```

```
## % latex table generated in R 4.2.1 by xtable 1.8-4 package
## % Thu Aug 10 10:47:22 2023
## \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) & -0.0006(0.001) & -0.0004(0.001) & -0.0006(0.000) & -0.0006(0.000) & -0.0005(0.00
##   Bias & 2. All except p(Z$|$A) &  0.3659(0.001) & -0.0012(0.002) & -0.0012(0.001) & -0.0480(0.000) &
##   Bias & 3. p(C), p(Z$|$A) & -0.0006(0.001) & -0.0004(0.001) & -0.0006(0.000) & -0.0006(0.000) & -0.0
##   Bias & 4. p(A$|$C), E(Y$|$Z,C) & -0.0006(0.001) & -0.0008(0.001) & -0.0012(0.001) & -0.0480(0.000)
##   EmpSE & 1. p(Z$|$A) &  0.0172(0.000) &  0.0164(0.000) &  0.0153(0.000) &  0.0153(0.000) &  0.0150(0
##   EmpSE & 2. All except p(Z$|$A) &  0.0180(0.000) &  0.0558(0.001) &  0.0472(0.001) &  0.0147(0.000)
##   EmpSE & 3. p(C), p(Z$|$A) &  0.0172(0.000) &  0.0164(0.000) &  0.0153(0.000) &  0.0153(0.000) &  0
##   EmpSE & 4. p(A$|$C), E(Y$|$Z,C) &  0.0171(0.000) &  0.0212(0.000) &  0.0472(0.001) &  0.0147(0.000)
##   MSE & 1. p(Z$|$A) &  0.0003(0.000) &  0.0003(0.000) &  0.0002(0.000) &  0.0002(0.000) &  0.0002(0.0
##   MSE & 2. All except p(Z$|$A) &  0.1342(0.000) &  0.0031(0.000) &  0.0022(0.000) &  0.0025(0.000) &
##   MSE & 3. p(C), p(Z$|$A) &  0.0003(0.000) &  0.0003(0.000) &  0.0002(0.000) &  0.0002(0.000) &  0.00
##   MSE & 4. p(A$|$C), E(Y$|$Z,C) &  0.0003(0.000) &  0.0005(0.000) &  0.0022(0.000) &  0.0025(0.000) &
##   ScEmpVar & 1. p(Z$|$A) &  14.8007(0.662) &  13.4974(0.604) &  11.7573(0.526) &  11.7573(0.526) &  1
##   ScEmpVar & 2. All except p(Z$|$A) &  16.1414(0.722) & 155.8208(6.969) & 111.4742(4.985) &  10.8023
##   ScEmpVar & 3. p(C), p(Z$|$A) &  14.8007(0.662) &  13.4974(0.604) &  11.7573(0.526) &  11.7573(0.526
##   ScEmpVar & 4. p(A$|$C), E(Y$|$Z,C) &  14.5580(0.651) &  22.5721(1.009) & 111.4742(4.985) &  10.8023
##   Bound &  & 14.765 & 22.5 & 18.71 & 11.864 & 17.995 & 11.15 \\
##    \hline
## \end{tabular}
## \end{table}
```

```r
rbind(result.bias.se.mse, result.sc.emp.var)%>%
        mutate(stat = recode(stat,  "bias" ="Bias", "empse" = "EmpSE", "mse" = "MSE", "ScEmpSE^2" = "Sc
```

```
##        stat      misspecification            BD             FD
## 1      Bias              1. p(Z|A) -0.0006(0.001) -0.0004(0.001)
## 2      Bias 2. All except p(Z|A)   0.3659(0.001) -0.0012(0.002)
## 3      Bias      3. p(C), p(Z|A)  -0.0006(0.001) -0.0004(0.001)
## 4      Bias  4. p(A|C), E(Y|Z,C)  -0.0006(0.001) -0.0008(0.001)
## 5     EmpSE              1. p(Z|A)  0.0172(0.000)  0.0164(0.000)
## 6     EmpSE 2. All except p(Z|A)   0.0180(0.000)  0.0558(0.001)
## 7     EmpSE      3. p(C), p(Z|A)   0.0172(0.000)  0.0164(0.000)
## 8     EmpSE  4. p(A|C), E(Y|Z,C)   0.0171(0.000)  0.0212(0.000)
## 9       MSE              1. p(Z|A)  0.0003(0.000)  0.0003(0.000)
## 10      MSE 2. All except p(Z|A)   0.1342(0.000)  0.0031(0.000)
## 11      MSE      3. p(C), p(Z|A)   0.0003(0.000)  0.0003(0.000)
## 12      MSE  4. p(A|C), E(Y|Z,C)   0.0003(0.000)  0.0005(0.000)
## 13 ScEmpVar              1. p(Z|A) 14.8007(0.662) 13.4974(0.604)
## 14 ScEmpVar 2. All except p(Z|A)  16.1414(0.722) 155.8208(6.969)
## 15 ScEmpVar      3. p(C), p(Z|A)  14.8007(0.662) 13.4974(0.604)
## 16 ScEmpVar  4. p(A|C), E(Y|Z,C)  14.5580(0.651) 22.5721(1.009)
## 17    Bound                              14.765           22.5
##               TD          BD TD           FD TD       BD FD TD
## 1  -0.0006(0.000) -0.0006(0.000)  -0.0005(0.000)  -0.0005(0.000)
## 2  -0.0012(0.001) -0.0480(0.000)  -0.0012(0.001)  -0.0480(0.000)
## 3  -0.0006(0.000) -0.0006(0.000)  -0.0919(0.000)  -0.0919(0.000)
## 4  -0.0012(0.001) -0.0480(0.000)  -0.0012(0.001)  -0.0480(0.000)
```

4

```
## 5    0.0153(0.000)   0.0153(0.000)   0.0150(0.000)   0.0150(0.000)
## 6    0.0472(0.001)   0.0147(0.000)   0.0472(0.001)   0.0147(0.000)
## 7    0.0153(0.000)   0.0153(0.000)   0.0145(0.000)   0.0145(0.000)
## 8    0.0472(0.001)   0.0147(0.000)   0.0472(0.001)   0.0147(0.000)
## 9    0.0002(0.000)   0.0002(0.000)   0.0002(0.000)   0.0002(0.000)
## 10   0.0022(0.000)   0.0025(0.000)   0.0022(0.000)   0.0025(0.000)
## 11   0.0002(0.000)   0.0002(0.000)   0.0086(0.000)   0.0086(0.000)
## 12   0.0022(0.000)   0.0025(0.000)   0.0022(0.000)   0.0025(0.000)
## 13  11.7573(0.526)  11.7573(0.526)  11.2103(0.501)  11.2103(0.501)
## 14 111.4742(4.985)  10.8023(0.483) 111.4742(4.985)  10.8023(0.483)
## 15  11.7573(0.526)  11.7573(0.526)  10.5289(0.471)  10.5289(0.471)
## 16 111.4742(4.985)  10.8023(0.483) 111.4742(4.985)  10.8023(0.483)
## 17          18.71          11.864          17.995           11.15
```