

Clustering standard errors and individual-level treatments

Mizuhiro Suzuki

2021-08-25

```
# Install and load packages -----
packages <- c(
  "tidyverse",
  "estimatr",
  "lfe",
  "fastmatch",
  "magrittr",
  "kableExtra",
  "fixest"
)

# Change to install = TRUE to install the required packages
pacman::p_load(packages, character.only = TRUE, install = FALSE)
```

Function for simulation

```
sim_func <- function(
  ind_fe,
  hetero_te,
  ind_assignment
) {

  set.seed(450166)
  # Number of people (population, not sample)
  NC <- 100000
  # Number of simulations
  num_sim <- 10000
  # Sample size
  sample_size = NC * 0.01

  # Treatment assignment probability
  mu <- 1/2 # mean assignment
  if (ind_assignment == TRUE) {
    max_sigma <- 1/4 # variability of assignment probability across individuals
  } else {
    max_sigma <- 0 # variability of assignment probability across individuals
  }

  # treatment probability for everyone in the population
  prob = runif(NC, min = mu - 2 * max_sigma, max = mu + 2 * max_sigma)
```

```

# individual fixed effects
eta_i = (ind_fe == TRUE) * rnorm(NC)
# treatment effect (can be individual-specific)
tau_i = (hetero_te == TRUE) * (c(rep(1, NC / 2), rep(-1, NC / 2)))

ind_df <- map_df(
  c(1, 2),
  ~ tibble(
    # individual index
    C = seq(NC),
    # random treatment
    W = rbinom(NC, 1, prob),
    # error term
    epsilon = rnorm(NC),
    # outcome
    Y = tau_i * W + eta_i + epsilon,
    # time
    period = .
  )
) %>%
  bind_rows() %>%
  select(Y, W, period, C)

pval_robust <- rep(0, num_sim)
pval_cluster <- rep(0, num_sim)
pval_robust_fe <- rep(0, num_sim)
pval_cluster_fe <- rep(0, num_sim)

for (i in seq(num_sim)) {

  sample_ind <- sample(seq(NC), sample_size, replace = FALSE)
  df_sample <- ind_df[ind_df$C %in% sample_ind,]

  res <- feols(
    Y ~ -1 + W, data = df_sample
  )
  pval_cluster[i] <- pvalue(res, cluster = "C")
  pval_robust[i] <- pvalue(res, se = "hetero")

  res <- feols(
    Y ~ -1 + W | C, data = df_sample
  )
  pval_cluster_fe[i] <- pvalue(res, cluster = "C")
  pval_robust_fe[i] <- pvalue(res, se = "hetero")

}

return(
  list(
    mean(pval_cluster < 0.05),
    mean(pval_robust < 0.05),
    mean(pval_cluster_fe < 0.05),
    mean(pval_robust_fe < 0.05)
  )
)

```

```

    )
  )
}

```

Baseline

```

res_1 <- sim_func(
  ind_fe = FALSE,
  hetero_te = FALSE,
  ind_assignment = FALSE
)

bind_cols(
  se_labels = c(
    "Cluster robust SE",
    "Heteroskedasticity robust SE",
    "Cluster robust SE, individual FE included in reg",
    "Heteroskedasticity robust SE, individual FE included in reg"
  ),
  share_p = unlist(res_1)
) %>%
  set_colnames(c(" ", "share of p-val < 0.05")) %>%
  kbl(caption = "Baseline", booktabs = T) %>%
  kable_styling(latex_options = "hold_position")

```

Table 1: Baseline

| | share of p-val < 0.05 |
|---|-----------------------|
| Cluster robust SE | 0.0519 |
| Heteroskedasticity robust SE | 0.0528 |
| Cluster robust SE, individual FE included in reg | 0.0486 |
| Heteroskedasticity robust SE, individual FE included in reg | 0.0486 |

Individual fixed effects (uncorrelated with treatments)

```

res_2 <- sim_func(
  ind_fe = TRUE,
  hetero_te = FALSE,
  ind_assignment = FALSE
)

bind_cols(
  se_labels = c(
    "Cluster robust SE",
    "Heteroskedasticity robust SE",
    "Cluster robust SE, individual FE included in reg",
    "Heteroskedasticity robust SE, individual FE included in reg"
  ),

```

```

share_p = unlist(res_2)
) %>%
  set_colnames(c(" ", "share of p-val < 0.05")) %>%
kbl(caption = "Individual fixed effects (uncorrelated with treatments)", booktabs = T) %>%
kable_styling(latex_options = "hold_position")

```

Table 2: Individual fixed effects (uncorrelated with treatments)

| | share of p-val < 0.05 |
|---|-----------------------|
| Cluster robust SE | 0.0478 |
| Heteroskedasticity robust SE | 0.0773 |
| Cluster robust SE, individual FE included in reg | 0.0486 |
| Heteroskedasticity robust SE, individual FE included in reg | 0.0486 |

Heterogeneous (= individual-specific) treatment effects

```

res_3 <- sim_func(
  ind_fe = FALSE,
  hetero_te = TRUE,
  ind_assignment = FALSE
)

bind_cols(
  se_labels = c(
    "Cluster robust SE",
    "Heteroskedasticity robust SE",
    "Cluster robust SE, individual FE included in reg",
    "Heteroskedasticity robust SE, individual FE included in reg"
  ),
  share_p = unlist(res_3)
) %>%
  set_colnames(c(" ", "share of p-val < 0.05")) %>%
kbl(caption = "Heterogeneous (= individual-specific) treatment effects", booktabs = T) %>%
kable_styling(latex_options = "hold_position")

```

Table 3: Heterogeneous (= individual-specific) treatment effects

| | share of p-val < 0.05 |
|---|-----------------------|
| Cluster robust SE | 0.0522 |
| Heteroskedasticity robust SE | 0.0823 |
| Cluster robust SE, individual FE included in reg | 0.0463 |
| Heteroskedasticity robust SE, individual FE included in reg | 0.0464 |

Within-individual serially correlated treatments

```

res_4 <- sim_func(
  ind_fe = FALSE,

```

```

hetero_te = FALSE,
ind_assignment = TRUE
)

bind_cols(
  se_labels = c(
    "Cluster robust SE",
    "Heteroskedasticity robust SE",
    "Cluster robust SE, individual FE included in reg",
    "Heteroskedasticity robust SE, individual FE included in reg"
  ),
  share_p = unlist(res_4)
) %>%
  set_colnames(c(" ", "share of p-val < 0.05")) %>%
kbl(caption = "Within-individual serially correlated treatments", booktabs = T) %>%
kable_styling(latex_options = "hold_position")

```

Table 4: Within-individual serially correlated treatments

| | share of p-val < 0.05 |
|---|-----------------------|
| Cluster robust SE | 0.0501 |
| Heteroskedasticity robust SE | 0.0506 |
| Cluster robust SE, individual FE included in reg | 0.0491 |
| Heteroskedasticity robust SE, individual FE included in reg | 0.0492 |