Integrate fast CRV3 for lm()

Alexander Fischer

4/23/2022

Example:

```
set.seed(98765)
library(clubSandwich)
library(clusterjack)

# few large clusters (around 10000 obs)
N <- 1000
N_G1 <-10
data <- fwildclusterboot:::create_data(
  N = N,
  N_G1 = N_G1,
  icc1 = 0.8,
  N_G2 = 10,
  icc2 = 0.8,
  numb_fe1 = 10,
  numb_fe2 = 10,
  seed = 12
)

lm_fit <- lm(proposition_vote ~ treatment  + log_income , data = data)

#clubSandwich::vcovCR(obj = lm_fit, cluster= data$group_id1, type = "CR3")

res <- clubSandwich:::vcov_CR(
  obj = lm_fit,
  cluster= data$group_id1,
  type = "CR3f"
)

res2 <- clusterjack::vcovJN(
  model = lm_fit,
  clustid= data$group_id1
)

all.equal(as.matrix(res), res2, check.attributes = FALSE)
```

```
## [1] TRUE
```

Comparison to CR3, check that all methods work

```
resCR3 <- clubSandwich:::vcov_CR(
  obj = lm_fit,
  cluster= data$group_id1,
  type = "CR3"
)
```

Confidence Interval:

```
clubSandwich::conf_int(
  obj = lm_fit,
  vcov = res,
  test = "Satterthwaite"
)
```

```
##        Coef. Estimate      SE d.f. Lower 95% CI Upper 95% CI
##  (Intercept)   1.0885 0.01695 7.23      1.04871      1.12838
##    treatment   0.0163 0.00986 8.97     -0.00598      0.03865
##   log_income  -0.0139 0.00218 6.30     -0.01918     -0.00864
```

```
clubSandwich::conf_int(
  obj = lm_fit,
  vcov = resCR3,
  test = "Satterthwaite"
)
```

```
##        Coef. Estimate     SE d.f. Lower 95% CI Upper 95% CI
##  (Intercept)   1.0885 0.0179 5.86      1.04457      1.13252
##    treatment   0.0163 0.0104 8.96     -0.00719      0.03986
##   log_income  -0.0139 0.0023 4.58     -0.01998     -0.00784
```

`coef_table()`:

```
clubSandwich::coef_test(
  obj = lm_fit,
  vcov = res,
  test = "Satterthwaite"
)
```

```
##        Coef. Estimate      SE t-stat d.f. (Satt) p-val (Satt) Sig.
##  (Intercept)   1.0885 0.01695  64.21        7.23       <0.001  ***
##    treatment   0.0163 0.00986   1.66        8.97        0.132
##   log_income  -0.0139 0.00218  -6.38        6.30       <0.001  ***
```

```
clubSandwich::coef_test(
  obj = lm_fit,
  vcov = resCR3,
  test = "Satterthwaite"
)
```

```
##        Coef. Estimate     SE t-stat d.f. (Satt) p-val (Satt) Sig.
##  (Intercept)   1.0885 0.0179  60.92        5.86       <0.001  ***
##    treatment   0.0163 0.0104   1.57        8.96       0.1506
##   log_income  -0.0139 0.0023  -6.06        4.58       0.0024   **
```

`Wald_test()`:

```
clubSandwich::Wald_test(
  obj = lm_fit,
  constraints = clubSandwich::constrain_zero(2:3, coef(lm_fit)),
  vcov = res
)
```

```
##  test Fstat df_num df_denom   p_val sig
##   HTZ    18      2     6.88 0.00185  **
```

```
clubSandwich::Wald_test(
  obj = lm_fit,
  constraints = clubSandwich::constrain_zero(2:3, coef(lm_fit)),
  vcov = resCR3
)
```

```
##  test Fstat df_num df_denom   p_val sig
##   HTZ  15.8      2     5.81 0.00446  **
```

Notes:

- cheat in small sample correction (not really using CR-adjustments.R)
- documentation - add note that `CR3f` only for `lm()`
- `vcovCR` still gets stuck somewhere, that's why I am using `clubSandwich::vcov_CR` here