Panel Data Models | R programming language

Preliminaries

Script Editor

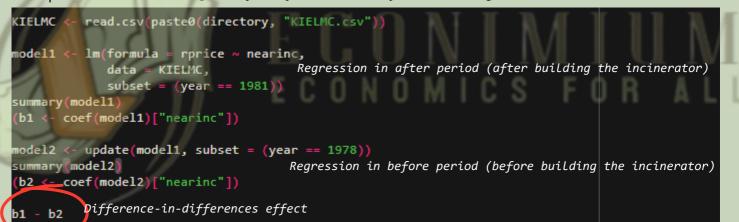


Difference-in-differences model

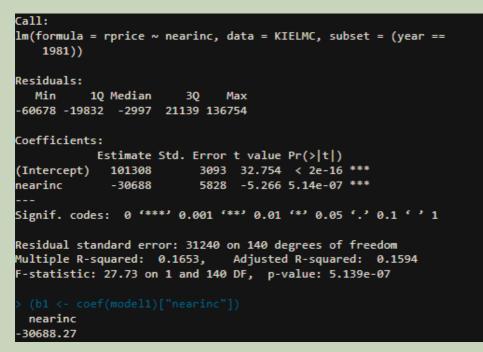
- Data: KIELMC.csv
- NOTE: Please read about DID models before attempting these commands.

• DID effect of building an incinerator on house prices.

Script Editor Deriving DID for after =0 and after =1 among treated and control units



Console



Panel Data Models | R programming language

Console ctd...

```
Call:
 lm(formula = rprice ~ nearinc, data = KIELMC, subset = (year ==
    1978))
 Residuals:
   Min
          1Q Median
                       ЗQ
                             Max
 -56517 -16605 -3193
                    8683 236307
 Coefficients:
           Estimate Std. Error t value Pr(>|t|)
                      2654 31.094 < 2e-16 ***
             82517
 (Intercept)
                         4745 -3.968 0.000105 ***
 nearinc
             -18824
 Signif. codes: 0 (**** 0.001 (*** 0.01 (** 0.05 (.' 0.1 ( ' 1
 Residual standard error: 29430 on 177 degrees of freedom
 Multiple R-squared: 0.08167, Adjusted R-squared: 0.07648
 F-statistic: 15.74 on 1 and 177 DF, p-value: 0.0001054
   nearinc
            Prices for houses near the incinerator were $11,864 lower than
  18824.37
            prices for houses far from the incinerator, after
            the incinerator was built.
  nearinc
  11863.9
                     Deriving DID for treated =0 and treated =1 among units before and after
Script Editor
model3 <- lm(rprice ~ y81, KIELMC, nearinc == 1)</pre>
summary(model3)
```

```
(b3 <- coef(model3)["y81"])
model4 <- update(model3, subset = (nearinc == 0))
summary(model4)
(b4 <- coef(model4)["y81"])
b3 - b4</pre>
```

Running this will give the same value for DID effect (-11863.9) in the console.

Instead of manually exploring DID as shown above, the same can be done altogether at once using a comprehensive DID regression which includes treated (dummy), after(dummy), and after*treated(interaction) as variables. DID effect is the coefficient on after*treated anbd will be the same as calculated above.

Script Editor

```
model5 <- lm(rprice ~ nearinc + y81 + y81nrinc, KIELMC)
summary(model5)
coef(model5)["y81nrinc"]</pre>
```

Panel Data Models | R programming language

Console

```
Call:
lm(formula = rprice ~ nearinc + y81 + y81nrinc, data = KIELMC)
Residuals:
  Min
          10 Median
                        30
                             Max
-60678 -17693 -3031 12483 236307
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
                      2727 30.260 < 2e-16 ***
(Intercept)
             82517
                         4875 -3.861 0.000137 ***
             -18824
nearinc
                                4.640 5.12e-06 ***
              18790
                         4050
y81
y81nrinc
             -11864
                         7457 -1.591 0.112595
Signif. codes: 0 (****) 0.001 (***) 0.01 (**) 0.05 (.' 0.1 (' 1
Residual standard error: 30240 on 317 degrees of freedom
Multiple R-squared: 0.1739, Adjusted R-squared: 0.1661
F-statistic: 22.25 on 3 and 317 DF, p-value: 4.224e-13
y81nrinc
-11863.9
                         More Panel Data Models: Preliminaries
```

- Example: Effect of training grants on firm scrap rate.
- Panel data where fcode is the cross sectional dimension and year is the time series dimension

Data: JTRAIN.csv

Script Editor

```
rm(list = ls())
directory <- "C:/Users/amalz/OneDrive/Desktop/"</pre>
PackageNames <- c("tidyverse", "stargazer", "magrittr", "haven", "plm")</pre>
for(i in PackageNames)
 if(!require(i, character.only = T))
    install.packages(i, dependencies = T)
    require(i, character.only = T)
JTRAIN <- read.csv(paste0(directory, 'JTRAIN.csv'))</pre>
JTRAIN %<>% filter(!is.na(lscrap)) <----- Drop missing observations for dependent variable
JTRAIN %<>% select(fcode, year, lscrap, tothrs, d88, d89, grant, grant_1)<----- Keep the variables needed</pre>
JTRAIN %>% group_by(fcode) %>% summarize(n = n()) <----- Number of observations by fcode</pre>
JTRAIN %>% group_by(year) %>% summarize(n = n())<----- Number of observations by year
JTRAIN
  select(lscrap, tothrs, d88, d89, grant, grant_1) %>%
  as.data.frame 🖇
  stargazer(type = "text", omit.summary.stat = "mean", digits = 2)
JTRAIN %>% group_by(fcode) %>%
  select(lscrap, tothrs, d88, d89, grant, grant_1) %>%
  summarize_all(mean) %>S
                                                                 {----- Overall variations
  as.data.frame
  select(-fcode) %
  stargazer(type = "text", digits = 2)
JTRAIN %>% group_by(fcode)
  select(lscrap, tothrs, d88, d89, grant, grant_1) %>%
                                                                {----- Within variations
  mutate_all(function(x) {x - mean(x)}) %
                                           % # d
  as.data.frame %>
  select(-fcode) %>9
  stargazer(type = "text", omit.summary.stat = "mean", digits =2)
```

Panel Data Models | R programming language

Pooled OLS, between, and first differences estimator

Script Editor

```
model_ols <- plm(formula = lscrap ~ tothrs + d88 + d89 + grant + grant_1,</pre>
                  data = JTRAIN,
                   index = c("fcode", "year"), # c(group index, time index)
                  model = "pooling")
summary(model_ols)
model_be <- update(model_ols, model = "between")</pre>
summary(model_be)
diff <- function(x) {x - dplyr::lag(x)}</pre>
JTRAIN %<>%
  group_by(fcode) %>%
  mutate(dlscrap = diff(lscrap),
          dtothrs = diff(tothrs),
dgrant = diff(grant)) %>%
  ungroup()
model fd <
            · lm(dlscrap ~
                          dtothrs + dgrant, JTRAIN)
summary(model_fd)
Console
Pooling Model
Call:
plm(formula = lscrap ~ tothrs + d88 + d89 + grant + grant_1,
   data = JTRAIN, model = "pooling", index = c("fcode", "year"))
Unbalanced Panel: n = 49, T = 2-3, N = 146
Residuals:
    Min.
           1st Qu.
                      Median
                              3rd Qu.
                                             Max.
-5.081736 -0.829602 -0.071229 1.078829 3.344532
Coefficients:
             Estimate Std. Error t-value Pr(>|t|)
(Intercept) 0.6625028 0.2231918 2.9683 0.003523 **
tothrs
           -0.0046484 0.0049508 -0.9389 0.349392
d88
           -0.2679208 0.3256257 -0.8228 0.412028
489
           -0.5128693 0.3658318 -1.4019 0.163150
grant
            0.3800988 0.3758131 1.0114 0.313568
grant_1
           0.0848627 0.4491363 0.1889 0.850408
____
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Total Sum of Squares:
                         310.97
Residual Sum of Squares: 302.34
               0.027761
R-Squared:
Adj. R-Squared: -0.0069617
F-statistic: 0.799506 on 5 and 140 DF, p-value: 0.5518
```

Panel Data Models | R programming language

Console Oneway (individual) effect Between Model Call: plm(formula = lscrap ~ tothrs + d88 + d89 + grant + grant_1, data = JTRAIN, model = "between", index = c("fcode", "year")) Unbalanced Panel: n = 49, T = 2-3, N = 146 Observations used in estimation: 49 Residuals: Min. 1st Qu. Median 3rd Qu. Max. -3.14987 -0.87378 -0.12629 0.95121 3.03382 Coefficients: (1 dropped because of singularities) Estimate Std. Error t-value Pr(>|t|) (Intercept) 1.2698730 1.7114534 0.7420 0.4620 -0.0077291 0.0107195 -0.7210 0.4747 tothrs d88 -3.1406816 5.1105974 -0.6145 0.5420 grant 2.4888824 1.8130559 1.3728 0.1768 grant_1 -1.1938015 1.7704187 -0.6743 0.5036 Total Sum of Squares: 93.159 Residual Sum of Squares: 88.943 0.045255 R-Squared: Adj. R-Squared: -0.04154 F-statistic: 0.521405 on 4 and 44 DF, p-value: 0.72044 summary(moc Call: lm(formula = dlscrap ~ dtothrs + dgrant, data = JTRAIN) Residuals: Min 10 Median Max 30 3.05113 -0.15989 0.05216 0.25951 2.60821 Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -0.210310 0.063422 -3.316 0.0013 ** 0.002612 -1.262 dtothrs -0.003297 0.2101 0.045589 0.115367 0.395 0.6936 dgrant Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.6034 on 93 degrees of freedom (66 observations deleted due to missingness) Multiple R-squared: 0.01949, Adjusted R-squared: -0.001597 -statistic: 0.9242 on 2 and 93 DF, p-value: 0.4004

Fixed effects within estimator

Script Editor

Panel Data Models | R programming language

Console

```
Oneway (individual) effect Within Model
Call:
plm(formula = lscrap ~ tothrs + d88 + d89 + grant + grant_1,
     data = JTRAIN, effect = "individual", model = "within", index = c("fcode",
          "year"))
Unbalanced Panel: n = 49, T = 2-3, N = 146
Residuals:
 Min. 1st Qu. Median 3rd Qu. Max.
-2.253927 -0.142083 -0.024063 0.159263 1.412912
Coefficients:
           Estimate Std. Error t-value Pr(>|t|)
tothrs -0.0047331 0.0029406 -1.6096 0.11092
d88
         -0.0747307
                      0.1212112 -0.6165 0.53907
         -0.2182447 0.1555527 -1.4030 0.16397
d89
         -0.1175244 0.1811571 -0.6487 0.51812
grant
grant_1 -0.4097214 0.2283540 -1.7942 0.07606
Signif. codes: 0 (**** 0.001 (*** 0.01 (** 0.05 (.' 0.1 ( ' 1
Total Sum of Squares:
                            31.521
Residual Sum of Squares: 24.311
R-Squared: 0.22873
Adj. R-Squared: -0.21559
F-statistic: 5.45676 on 5 and 92 DF, p-value: 0.00019242
                                                                                                      410593
                                               410565
                                                            410566
                                                                       410567
     410523
                   410538
                                 410563
                                                                                       410577
              0.45337572 1.29699872 0.96969656 1.14588496 -1.19438345 -0.05871249 -0.56425581
 3.27833527
      410596
                   410606
 1.31942131 -1.30365914
Estimate Std. Error t-value Pr(>|t|)
410523 -3.2783353 0.3326591 -9.8549 4.609e-16
410538 0.4533757 0.3082121 1.4710 0.1447070
                                                      ***
410563 1.2969987 0.3100223 4.1836 6.549e-05 ***
410565 0.9696966 0.3076827 3.1516 0.0021919 **
410566 1.1458850 0.3076683 3.7244 0.0003373 ***
410567 -1.1943834 0.3076827 -3.8819 0.0001948 ***
410577 -0.0587125 0.3076418 -0.1908 0.8490661
410593 -0.5642558 0.3078266 -1.8330 0.0700321 .
410596 1.3194213 0.3086255 4.2752 4.659e-05 ***
410606 -1.3036591 0.3095618 -4.2113 5.910e-05 ***
410626 -0.6773954 0.3134209 -2.1613 0.0332689 *
410665 -3.5231088 0.3083375 -11.4261 < 2.2e-16 ***
418011 1.1072240 0.3128219 3.5395 0.0006310 ***
418021 -0.3096219 0.3095171 -1.0003 0.3197706
418035 1.1477166 0.3093131 3.7105 0.0003538 ***
418045 0.4759869 0.3696770 1.2876 0.2011234
418051 -1.0252676 0.3100117 -3.3072 0.0013454 **
418054 -0.1049205 0.3095171 -0.3390 0.7353962
418065 -0.1138331 0.3091246 -0.3682 0.7135377
418076 -0.7083632 0.3300305 -2.1464 0.0344758 *
418083 0.2771049 0.3741704 0.7406 0.4608312
                     0.3101662
        0.0799539
418091
                                   0.2578 0.7971542
418097 0.0068739 0.3091242
                                 0.0222 0.9823073
418107 -0.7154665 0.3175040 -2.2534 0.0266090 *
Signif. codes: 0 (**** 0.001 (*** 0.01 (** 0.05 (.' 0.1 ( ' 1
 Statistic N Mean St. Dev. Min Max
ai
           49 0.002 1.404 -3.523 2.722
```

Panel Data Models | R programming language

Dummy variables regression

Script Editor

Script Editor

• factor(fcode) creates one dummy variable for each fcode. Results will be similar to fixed effects estimator shown above.

Random effects estimator

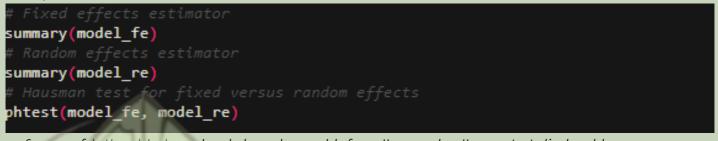
model_re <- plm(formula = lscrap ~ tothrs + d88 + d89 + grant + grant_1,</pre> data = JTRAIN, index = c("fcode", "year"), # c(group index, time index) model = "random", random.method = "walhus") summary(model_re (sigma2_e <- model_re\$ercomp\$sigma2["idios"])</pre> (sigma2_u <- model_re\$ercomp\$sigma2["id"]) <- 1 - sqrt(sigma2_e / (sigma2_e + 3*sigma2_u))) (theta Console Oneway (individual) effect Random Effect Model (Wallace-Hussain's transformation) Call: "walhus", Unbalanced Panel: n = 49, T = 2-3, N = 146 Effects: var std.dev share idiosyncratic 0.2553 0.5053 0.118 individual 1.9172 1.3846 0.882 theta: Min. 1st Qu. Median Mean 3rd Qu. Max. 0.7501 0.7938 0.7938 0.7932 0.7938 0.7938 Residuals: Min. 1st Qu. Median Mean 3rd Qu. Max. -2.51427 -0.19917 -0.00172 -0.00019 0.27108 1.59999 Coefficients: Estimate Std. Error z-value Pr(>|z|) Intercept) 0.6638756 0.2166307 3.0646 0.00218 ** tothrs -0.0047381 0.0028841 -1.6897 0.09108 . d88 -0.0915398 0.1194667 -0.7662 0.44354 d89 -0.2483872 0.1517170 -1.6372 0.10159 -0.0741720 0.1751030 -0.4236 0.67186 -0.3546065 0.2203972 -1.6089 0.10763 grant grant_1 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Total Sum of Squares: 43.404 Residual Sum of Squares: 36.353 R-Squared: 0.16246 Adj. R-Squared: 0.13255 Chisq: 27.1435 on 5 DF, p-value: 5.3487e-05 idios 0.2553429 id 1,917162 idios .7938235

Panel Data Models | R programming language

Hausman test for fixed versus random effects

- The Hausman test is used to decide whether to use fixed effects or random effects.
- H0: FE coefficients are not significantly different from the RE coefficients
- Ha: FE coefficients are significantly different from the RE coefficients
- If the Hausman test statistic is insignificant, use RE estimator because it is efficient
- If the Hausman test statistic is significant, use FE estimator because it is consistent

Script Editor



Summary of both models have already been showcased before. Hence, only Hausman test displayed here. Console

