Model for pooled data

Henri Makika

Julho 04, 2019

Data Agriculture climate

```
library(readxl)
library(stargazer)

Data_AgricultureClimate <- read_excel("~/Videos/Inverno 2019/Aula 4/Data_AgricultureClimate.xls")

head(Data_AgricultureClimate)
```

```
## # A tibble: 6 x 7
##     ano periodo   vtotal areatotal tempano precano   munic
##   <dbl>   <dbl>    <dbl>     <dbl>   <dbl>   <dbl>   <dbl>
## 1  1990       0 1264259.   1388970    22.9   1316. 3500105
## 2  1991       0  737240.   1235817    23.2   1415. 3500105
## 3  1992       0 1381728.   1059597    22.7   1390. 3500105
## 4  1993       0 1547279.    941193    23.2   1320. 3500105
## 5  1994       0 1383900.    991584    23.6   1160. 3500105
## 6  1995       0 1632009.    972081    23.6   1492. 3500105
```

```
summary(Data_AgricultureClimate)
```

```
##       ano          periodo         vtotal           areatotal
##  Min.   :1990   Min.   :0.00   Min.   :       4   Min.   :       0
##  1st Qu.:1996   1st Qu.:0.00   1st Qu.:  591117   1st Qu.:  227997
##  Median :2002   Median :0.00   Median : 1598384   Median :  649044
##  Mean   :2002   Mean   :0.36   Mean   : 2851582   Mean   : 1191138
##  3rd Qu.:2008   3rd Qu.:1.00   3rd Qu.: 3598698   3rd Qu.: 1487277
##  Max.   :2014   Max.   :1.00   Max.   :45904552   Max.   :14584086
##                                NA's   :1552       NA's   :375
##     tempano         precano          munic
##  Min.   :14.34   Min.   : 155.3   Min.   :3500105
##  1st Qu.:20.53   1st Qu.:1226.1   1st Qu.:3514576
##  Median :21.59   Median :1373.8   Median :3528552
##  Mean   :21.79   Mean   :1406.4   Mean   :3528625
##  3rd Qu.:23.27   3rd Qu.:1542.8   3rd Qu.:3542753
##  Max.   :25.19   Max.   :4692.0   Max.   :3557204
##
```

Regression 1

```
loglinear = lm(log(vtotal) ~ log(areatotal) + (tempano) + (precano),
               data = Data_AgricultureClimate)
```

Regression 2

```
loglinear2 = lm(log(vtotal) ~ log(areatotal) + (tempano) + (precano) + periodo,
               data = Data_AgricultureClimate)
```

Mudança no coefficients

```
int_areatotal = Data_AgricultureClimate$areatotal*Data_AgricultureClimate$periodo

loglinear3 = lm(log(vtotal) ~ log(areatotal) + (tempano) + (precano) + periodo
                + (int_areatotal),
               data = Data_AgricultureClimate)

stargazer(loglinear, loglinear2, loglinear3, type = "text", digits = 5,
          column.labels = c("", "", ""), keep.stat = c('n','rsq','adj.rsq','f'),
          out = "mrd.txt")
```

```
##
## ==========================================================================================================
##                                                  Dependent variable:
##                 ---------------------------------------------------------------------------------------
##                                                      log(vtotal)
##
##                              (1)                          (2)                          (3)
## ----------------------------------------------------------------------------------------------------------
## log(areatotal)           0.95245***                   0.93920***                   0.94798***
##                           (0.00375)                    (0.00360)                    (0.00401)
##
## tempano                  -0.01594***                  -0.02095***                  -0.02201***
##                           (0.00336)                    (0.00320)                    (0.00321)
##
## precano                   0.00004*                     0.00002                      0.00002
##                           (0.00002)                    (0.00002)                    (0.00002)
##
## periodo                                                0.37914***                   0.41733***
##                                                         (0.01069)                    (0.01316)
##
## int_areatotal                                                                      -0.0000000***
##                                                                                     (0.00000)
##
## Constant                 1.70869***                   1.88977***                   1.79947***
##                           (0.09271)                    (0.08854)                    (0.09031)
##
## ----------------------------------------------------------------------------------------------------------
## Observations               12,588                       12,588                       12,588
## R2                         0.83962                      0.85421                      0.85449
## Adjusted R2                0.83958                      0.85416                      0.85444
## F Statistic    21,960.18000*** (df = 3; 12584) 18,431.20000*** (df = 4; 12583) 14,777.63000*** (df =
## ==========================================================================================================
## Note:                                                                          *p<0.1; **p<0.05;
```

Data Panel

L'objectif des données de panel est donc de contrôler les variables que la regréssion linéaire ne saura pas réaliser. Plusieures variables pour éviter les biais d'informations dans la regréssion. On peut donc créer une variable binaire pour controler les scénarios.

```
library(plm)

panelregre = plm(log(vtotal) ~ log(areatotal) + tempano + precano, index = c("munic", "ano"),
                 model = "within", data = Data_AgricultureClimate)

stargazer(panelregre, type = "text", digits = 5, column.labels = c(""),
          keep.stat = NULL, out = "mrd.txt")
```

```
##
## ===============================================
##                       Dependent variable:
##                   ----------------------------
##                            log(vtotal)
##
## -----------------------------------------------
## log(areatotal)             0.93367***
##                            (0.00964)
##
## tempano                    0.18908***
##                            (0.00733)
##
## precano                   -0.00006***
##                            (0.00002)
##
## -----------------------------------------------
## Observations                 12,588
## R2                           0.44493
## Adjusted R2                  0.41951
## F Statistic    3,215.85900*** (df = 3; 12036)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
panelregre2 = plm(log(vtotal) ~ log(areatotal) + tempano + precano, index = c("munic", "ano"),
                  effect = "twoways", model = "within", data = Data_AgricultureClimate)

stargazer(panelregre, panelregre2, type = "text", digits = 5, column.labels = c("", ""),
          keep.stat = NULL, out = "mrd.txt")
```

```
##
## ============================================================================
##                                      Dependent variable:
##                   ---------------------------------------------------------
##                                           log(vtotal)
##
##                               (1)                            (2)
## ----------------------------------------------------------------------------
## log(areatotal)             0.93367***                     0.86365***
##                            (0.00964)                      (0.00763)
##
```

```
## tempano                  0.18908***                    0.00611
##                          (0.00733)                    (0.00700)
##
## precano                 -0.00006***                    0.00003*
##                          (0.00002)                    (0.00002)
##
## ------------------------------------------------------------------------
## Observations               12,588                       12,588
## R2                        0.44493                      0.51927
## Adjusted R2               0.41951                      0.49626
## F Statistic    3,215.85900*** (df = 3; 12036) 4,325.06500*** (df = 3; 12012)
## ========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

Random Effects and Hausman

L'avantage des données en panel ce que le degré de liberté augmente $p(n-k) = s.e$, mais on perd l'efficience avec $E(\hat{\beta}) = 0$. En donnée de panel, on inclut *dummy* pour capter toute les informations constantes au long du temps. C'est à dire on contrôle toute les variables omis. (voir aula 4, p. 18, slides). n égal recette = taille du modèle = amostra et k égal numéro de variables.

Dans le modèle *Random Effects* e_{it} mesure la variance entre la ligne de regréssion et C_i qui est la variable *dummy*, mesure la variance générale entre toute le ligne (aula 5, p. 3, slides).

```
random1 = plm(log(vtotal) ~ log(areatotal) + tempano, data = Data_AgricultureClimate,
              index = c("munic", "ano"), model = "random")

random2 = plm(log(vtotal) ~ log(areatotal) + tempano + as.factor(ano),
              data = Data_AgricultureClimate, index = c("munic", "ano"), model = "random")

stargazer(random1, random2, type = "text", digits = 5, column.labels = c("", ""),
          keep.stat = NULL, out = "mrd.txt")
```

```
##
## ============================================================================
##                                      Dependent variable:
##                     ---------------------------------------------------------
##                                          log(vtotal)
##
##                                (1)                          (2)
## ----------------------------------------------------------------------------
## log(areatotal)              0.92678***                   0.88993***
##                             (0.00712)                    (0.00604)
##
## tempano                     0.12371***                   -0.00297
##                             (0.00606)                    (0.00582)
##
## as.factor(ano)1991                                       -0.22795***
##                                                          (0.02057)
##
## as.factor(ano)1992                                       0.45106***
##                                                          (0.02060)
##
## as.factor(ano)1993                                       0.61571***
```

```
##                                                    (0.02069)
##
## as.factor(ano)1994                                0.59632***
##                                                    (0.02091)
##
## as.factor(ano)1995                                0.54243***
##                                                    (0.02098)
##
## as.factor(ano)1996                                0.57063***
##                                                    (0.02101)
##
## as.factor(ano)1997                                0.53620***
##                                                    (0.02093)
##
## as.factor(ano)1998                                0.62035***
##                                                    (0.02101)
##
## as.factor(ano)1999                                0.55050***
##                                                    (0.02088)
##
## as.factor(ano)2000                                0.65207***
##                                                    (0.02076)
##
## as.factor(ano)2001                                0.85050***
##                                                    (0.02120)
##
## as.factor(ano)2002                                0.95138***
##                                                    (0.02233)
##
## as.factor(ano)2003                                0.82917***
##                                                    (0.02115)
##
## as.factor(ano)2004                                0.81623***
##                                                    (0.02088)
##
## as.factor(ano)2005                                0.91050***
##                                                    (0.02121)
##
## as.factor(ano)2006                                1.05524***
##                                                    (0.02120)
##
## as.factor(ano)2007                                1.05637***
##                                                    (0.02173)
##
## as.factor(ano)2008                                0.90107***
##                                                    (0.02098)
##
## as.factor(ano)2009                                0.83407***
##                                                    (0.02128)
##
## as.factor(ano)2010                                1.00665***
##                                                    (0.02117)
##
## as.factor(ano)2011                                1.05223***
```

```
##                                                                  (0.02107)
##
## as.factor(ano)2012                                              0.98717***
##                                                                  (0.02149)
##
## as.factor(ano)2013                                              0.91078***
##                                                                  (0.02099)
##
## as.factor(ano)2014                                              0.84232***
##                                                                  (0.02188)
##
## Constant                          -0.96151***                   1.59258***
##                                    (0.16036)                    (0.14846)
##
## ------------------------------------------------------------------------------
## Observations                        12,588                        12,588
## R2                                 0.58845                       0.74666
## Adjusted R2                        0.58838                       0.74614
## F Statistic        8,993.91100*** (df = 2; 12585) 1,423.47200*** (df = 26; 12561)
## ==============================================================================
## Note:                                          *p<0.1; **p<0.05; ***p<0.01
```

Hausman test

Ce test compare la consistance de différents estimateurs alternatifs. Grâce à cela, ce test permet de vérifier si le modèle économétrique est approprié au cas traité par l'économiste.

```
phtest(panelregre, random1)
```

```
##
##  Hausman Test
##
## data:  log(vtotal) ~ log(areatotal) + tempano + precano
## chisq = 276.3, df = 2, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

```
phtest(panelregre2, random2)
```

```
##
##  Hausman Test
##
## data:  log(vtotal) ~ log(areatotal) + tempano + precano
## chisq = 53.106, df = 2, p-value = 2.939e-12
## alternative hypothesis: one model is inconsistent
```

Dans le deux cas, on accepte H_0.