Panel Data Estimation

Zhentao Shi

November 22, 2017

This document demonstrates panel data estimation methods.

Dataset

```
g0 <- read.csv("http://www.nber.org/nberces/nberces5809/naics5809.csv")
```

The data comes from NBER-CES Manufacturing Industry Database. The data size is about 4M. Downloading would take up to a few minutes if the network is slow.

The dataset contains annual information of 473 USA industries during 1958 to 2009. To have some idea what a panel data looks like, we display the first 100 rows and 10 columns.

```
g0[1:80, 1:10]
```

```
##     naics year  emp   pay prode prodh prodw   vship matcost   vadd
## 1  311111 1958 18.0  81.3  12.0  25.7  49.8  1042.4   752.4  266.9
## 2  311111 1959 17.9  82.5  11.8  25.5  49.4  1051.0   758.9  268.7
## 3  311111 1960 17.7  84.8  11.7  25.4  50.0  1050.2   752.8  269.9
## 4  311111 1961 17.5  87.4  11.5  25.4  51.4  1119.7   803.6  287.8
## 5  311111 1962 17.6  90.2  11.5  25.2  52.1  1175.7   853.3  294.5
## 6  311111 1963 17.1  89.8  11.0  23.9  52.1  1249.1   893.6  328.7
## 7  311111 1964 16.6  90.8  10.6  23.5  52.2  1245.6   890.2  326.8
## 8  311111 1965 16.0  90.8  10.2  22.7  51.8  1283.5   928.1  324.7
## 9  311111 1966 16.1  96.1  10.2  22.6  53.9  1428.8  1049.9  344.8
## 10 311111 1967 16.7 105.0  11.0  23.9  61.3  1544.1  1101.6  410.0
## 11 311111 1968 16.6 109.6  10.9  23.5  64.0  1532.6  1070.9  426.6
## 12 311111 1969 17.3 119.5  11.4  24.4  70.6  1638.0  1134.0  470.2
## 13 311111 1970 17.9 130.0  12.1  25.3  79.8  1788.8  1243.2  515.1
## 14 311111 1971 17.0 132.8  11.5  23.8  81.5  1870.5  1309.3  518.0
## 15 311111 1972 12.5 121.6   9.7  21.2  88.1  1260.4   682.9  572.8
## 16 311111 1973 13.7 143.6  10.7  24.1 103.7  1726.1  1041.1  685.4
## 17 311111 1974 13.9 154.1  10.8  22.5 110.8  1869.3  1088.9  771.5
## 18 311111 1975 14.3 172.1  10.9  23.2 123.5  2091.2  1183.5  882.0
## 19 311111 1976 14.4 191.5  10.9  23.4 134.5  2405.5  1318.8 1061.2
## 20 311111 1977 15.5 220.9  11.8  24.5 155.8  2775.1  1458.8 1299.9
## 21 311111 1978 16.3 246.1  12.5  26.3 172.8  2968.3  1499.6 1452.5
## 22 311111 1979 16.3 261.0  12.1  24.9 175.5  2903.9  1573.1 1303.7
## 23 311111 1980 16.7 282.2  12.1  24.2 188.9  3288.7  1646.5 1626.4
## 24 311111 1981 15.2 283.6  10.7  22.1 187.7  3416.1  1655.2 1741.0
## 25 311111 1982 15.2 306.1  11.3  22.7 206.3  3957.8  1836.7 2129.0
## 26 311111 1983 15.0 308.2  11.1  22.1 209.7  4293.8  1997.1 2268.0
```

```
## 27 311111 1984 15.2 329.9  11.3  22.0 218.5  4417.7  1973.6 2451.1
## 28 311111 1985 14.6 347.4  10.6  20.8 223.5  4770.5  1932.6 2868.4
## 29 311111 1986 14.2 371.6  10.3  20.9 244.2  4925.2  2029.3 2933.5
## 30 311111 1987 13.4 365.8   9.9  20.5 244.2  5069.3  2296.8 2741.5
## 31 311111 1988 13.7 384.6  10.1  20.9 254.4  5956.4  2911.0 3088.4
## 32 311111 1989 13.2 395.1   9.8  20.1 263.6  6703.3  3149.2 3577.6
## 33 311111 1990 12.9 395.4   9.5  20.0 262.3  7015.0  3210.5 3842.2
## 34 311111 1991 12.8 405.0   9.6  20.6 271.4  7097.4  3467.3 3619.8
## 35 311111 1992 13.8 455.6  10.5  22.4 301.5  7023.9  3295.5 3729.9
## 36 311111 1993 14.1 477.6  10.7  23.5 323.5  7245.3  3591.7 3643.0
## 37 311111 1994 13.3 454.4   9.8  22.1 302.9  6938.2  3465.2 3477.5
## 38 311111 1995 13.4 464.1   9.6  22.0 301.0  7253.0  3961.9 3279.1
## 39 311111 1996 13.3 484.5  10.1  23.3 330.5  7572.2  4113.4 3496.6
## 40 311111 1997 14.0 502.2  10.6  23.4 345.4  8688.2  4402.1 4307.1
## 41 311111 1998 14.1 528.1  10.8  24.0 370.1  8967.1  4548.6 4396.8
## 42 311111 1999 14.2 557.5  11.0  24.0 380.3  8559.6  4324.9 4249.7
## 43 311111 2000 14.8 583.8  11.5  25.1 390.0  8751.4  4471.4 4283.0
## 44 311111 2001 14.1 568.0  10.7  22.7 382.0  9734.9  4576.0 5153.9
## 45 311111 2002 14.5 635.0  11.0  23.4 436.7 10662.2  4786.7 5924.7
## 46 311111 2003 14.2 627.7  11.1  23.8 436.7 11006.9  4713.7 6270.0
## 47 311111 2004 13.0 618.4  10.1  22.1 438.8 12127.9  5405.7 6722.6
## 48 311111 2005 14.5 680.8  11.2  24.5 484.6 13169.9  5829.4 7355.0
## 49 311111 2006 14.6 697.4  11.5  24.7 499.7 13596.6  5805.3 7804.2
## 50 311111 2007 17.0 792.9  13.1  28.3 553.0 14390.2  6917.4 7523.4
## 51 311111 2008 16.7 809.1  12.8  27.6 561.6 16997.8  7923.0 9077.1
## 52 311111 2009 17.1 882.0  13.0  27.7 620.5 19691.0  9776.6 9893.6
## 53 311119 1958 39.2 170.6  25.9  55.5 101.8  2194.1  1690.9  531.5
## 54 311119 1959 38.9 173.2  25.4  54.9 101.1  2212.1  1705.5  534.8
## 55 311119 1960 38.6 178.0  25.1  54.5 102.5  2210.3  1691.7  537.3
## 56 311119 1961 38.2 183.4  24.8  54.8 105.4  2356.7  1805.9  573.1
## 57 311119 1962 38.5 189.2  24.8  54.1 106.8  2474.4  1917.9  586.4
## 58 311119 1963 37.4 188.4  23.5  51.6 106.7  2628.7  2008.3  654.4
## 59 311119 1964 36.1 190.4  22.9  50.6 107.0  2621.7  2000.8  650.7
## 60 311119 1965 35.1 190.6  21.8  48.9 106.1  2701.3  2086.0  646.3
## 61 311119 1966 35.2 201.8  21.7  48.7 110.3  3007.0  2359.8  686.5
## 62 311119 1967 36.5 220.4  23.5  51.2 125.5  3250.0  2475.8  816.1
## 63 311119 1968 36.4 230.1  23.4  50.7 131.1  3225.6  2406.7  849.3
## 64 311119 1969 38.0 250.9  24.4  52.6 144.8  3447.5  2548.9  936.4
## 65 311119 1970 39.1 272.9  26.0  54.4 163.4  3764.6  2794.2 1025.8
## 66 311119 1971 37.2 278.8  24.8  51.2 167.0  3936.8  2942.5 1031.5
## 67 311119 1972 45.6 353.2  28.9  63.2 196.3  5174.1  4056.9 1140.5
## 68 311119 1973 42.7 366.6  26.4  60.6 200.5  6964.9  5666.1 1355.6
## 69 311119 1974 45.2 410.8  27.9  59.8 219.3  7934.2  6508.6 1460.5
## 70 311119 1975 44.4 427.8  27.7  59.4 225.8  7493.4  6060.1 1446.2
## 71 311119 1976 43.4 458.5  26.8  57.9 239.5  8210.2  6588.4 1668.5
## 72 311119 1977 41.1 486.8  24.7  51.2 253.5  9090.5  7474.6 1636.3
## 73 311119 1978 42.8 546.6  25.1  52.3 275.2  9213.5  7516.8 1726.3
## 74 311119 1979 41.4 552.4  25.5  54.1 296.2 10543.0  8501.5 2084.1
```

```
## 75 311119 1980 41.2 600.3  25.1  53.2 318.6 11120.4  9022.1 2135.3
## 76 311119 1981 39.6 606.6  23.4  51.2 325.3 11718.5  9595.9 2130.7
## 77 311119 1982 39.6 649.9  23.1  48.0 346.0 11732.3  9391.7 2338.8
## 78 311119 1983 38.3 659.8  22.7  46.3 352.7 12237.4  9843.8 2416.2
## 79 311119 1984 36.5 655.1  20.8  42.3 343.4 12387.8  9808.4 2559.0
## 80 311119 1985 34.0 633.0  18.9  38.5 330.0 10936.0  8413.6 2479.3
```

Estimation

`install.packages("plm")` if you use the package plm for the first time. An introduction can be found here. Load the package.

```
library(plm)
```

It is very important to explicitly define which column is the cross-sectional dimension and which one is the time dimension.

```
g <- pdata.frame( g0, index = c("naics", "year") )
```

Now we are ready for estimation. Suppose we are interested in, for the purpose of demonstration, a regression with a dependent variable **emp** and explanatory variables **investment** and **capital**. We write down the formula as it will be used repeatedly.

```
equation <- emp~invest+cap
```

OLS and Pooled OLS

Nothing prevents from running an OLS.

```
g.ols <- lm(equation, data=g)
summary(g.ols)
```

```
##
## Call:
## lm(formula = equation, data = g)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -364.61  -17.88   -9.57    6.42  416.23
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.430e+01  2.655e-01  91.509  < 2e-16 ***
## invest      -5.393e-03  8.766e-04  -6.152 7.77e-10 ***
## cap          4.120e-03  6.341e-05  64.971  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.89 on 24164 degrees of freedom
```

```
##   (429 observations deleted due to missingness)
## Multiple R-squared:  0.2927, Adjusted R-squared:  0.2926
## F-statistic:  5000 on 2 and 24164 DF,  p-value: < 2.2e-16
```

The OLS coefficient estimates are exactly the same as the pooled OLS. The only difference in the summary is that the later shows the panel structure of the data.

```
g.pool <- plm(equation,data=g,model="pooling")
summary(g.pool)
```

```
## Pooling Model
##
## Call:
## plm(formula = equation, data = g, model = "pooling")
##
## Unbalanced Panel: n=473, T=13-52, N=24167
##
## Residuals :
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -364.6116  -17.8760   -9.5675    6.4165  416.2347
##
## Coefficients :
##                Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept)  2.4296e+01  2.6550e-01  91.509 < 2.2e-16 ***
## invest      -5.3929e-03  8.7660e-04  -6.152 7.771e-10 ***
## cap          4.1200e-03  6.3413e-05  64.971 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    49051000
## Residual Sum of Squares: 34694000
## R-Squared:      0.2927
## Adj. R-Squared: 0.29264
## F-statistic: 4999.87 on 2 and 24164 DF, p-value: < 2.22e-16
```

Random Effect and Fixed Effect

The coefficient estimates differ in the random effect and the fixed effect.

```
g.re <- plm(equation, data=g, model="random")
summary(g.re)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = equation, data = g, model = "random")
##
## Unbalanced Panel: n=473, T=13-52, N=24167
```

```
##
## Effects:
##                   var std.dev share
## idiosyncratic  335.71   18.32  0.24
## individual    1061.41   32.58  0.76
## theta  :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.8459  0.9222  0.9222  0.9218  0.9222  0.9222
##
## Residuals :
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -191.390   -4.824   -1.260   -0.007    3.535  242.581
##
## Coefficients :
##                Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept)  2.9718e+01  1.5116e+00 19.6597 < 2.2e-16 ***
## invest      -4.2847e-03  5.5075e-04 -7.7798 7.553e-15 ***
## cap          2.1374e-03  6.8567e-05 31.1729 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    8671600
## Residual Sum of Squares: 8141400
## R-Squared:      0.061142
## Adj. R-Squared: 0.061064
## F-statistic: 786.825 on 2 and 24164 DF, p-value: < 2.22e-16
```

```
g.fe <- plm(equation, data=g, model="within")
summary(g.fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = equation, data = g, model = "within")
##
## Unbalanced Panel: n=473, T=13-52, N=24167
##
## Residuals :
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -212.735344   -3.948681   -0.020028    3.965494  233.204238
##
## Coefficients :
##           Estimate  Std. Error t-value  Pr(>|t|)
## invest -3.8758e-03  5.5301e-04 -7.0086 2.471e-12 ***
## cap     2.0277e-03  6.9677e-05 29.1009 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Total Sum of Squares:    8420700
## Residual Sum of Squares: 7953600
## R-Squared:      0.055468
## Adj. R-Squared: 0.036571
## F-statistic: 695.667 on 2 and 23692 DF, p-value: < 2.22e-16
```

Which model is preferred? The Hausman test favors the fixed-effect model.

```
phtest(g.re, g.fe)
```

```
##
##  Hausman Test
##
## data:  equation
## chisq = 65.835, df = 2, p-value = 5.059e-15
## alternative hypothesis: one model is inconsistent
```