

# Panel data

Ryuya Ko

12/6/2019

## パネルデータ分析

- パネルデータ: クロスセクションデータを時系列に集計したもの。個別の観察単位に関して、複数期間に渡って観察・サンプリングしたもの  
したがって、モデルは次のように表される:

$$y_{it} = x'_{it}\beta + \varepsilon_{it} \text{ for } i = 1, \dots, n, \ t = 1, \dots, T.$$

そして、誤差項  $\varepsilon_{it}$  が次のように分解できると仮定する:

$$\varepsilon_{it} = \alpha_i + v_{it}.$$

ここで、 $\alpha_i$  は時間を通じて一定の個別効果であり、 $v_{it}$  は真の意味での誤差項であり、 $E[v_{it}|x_{i1}, \dots, x_{iT}] = 0$  が成り立つ。

もし個別効果を  $\alpha_i = z'_i\gamma$  などとして観察できるか、 $\alpha_i = \alpha$  とできるならば、これは通常の OLS として扱うことができる (Pooled Regression)。

そうではない場合、通常では以下の 2 つのモデルが考えられる:

- 固定効果モデル (Fixed Effect Model): 観測されない個別効果  $\alpha_i$  が存在し、 $E[\alpha_i|x_{i1}, \dots, x_{iT}] \neq 0$  である。
- ランダム効果モデル (Random Effect Model): 観測されない個別効果  $u_i$  が存在し、 $\alpha_i|x_{i1}, \dots, x_{iT} \sim (0, \sigma_u^2)$  である。

## 固定効果モデル

### Within Estimator

個別効果は時間を通じて一定であることから、各  $i$  について時間平均の偏差を求めることによって個別効果を除外できる:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)' \beta + v_{it} - \bar{v}_i$$

$E[v_{it}|x_{i1}, \dots, x_{iT}] = 0$  の仮定より、 $y_{it} - \bar{y}_i$  を  $(x_{it} - \bar{x}_i)$  に回帰することで BLUE な推定量を得られる。これを FE estimator と呼ぶ。

FE estimator は次のように表され、Within estimator と呼ばれる：

$$b_{FE} = [S_{xx}^{within}]^{-1} S_{xy}^{within}$$

$$S_{xx}^{within} = \sum_{i=1}^n \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'$$

$$S_{xy}^{within} = \sum_{i=1}^n \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)$$

## ダミー変数による表現

固定効果をダミー変数を用いて表すことによって、FE モデルは Least Squares Dummy Variable model と呼ばれる。(板書)

## R による実装

```
# setwd('~/.econ/subsemi/introductory_econometrics/')
library(dplyr); library(wooldridge);

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(plm); library(stargazer)

##
## Attaching package: 'plm'
```

```
## The following objects are masked from 'package:dplyr':
##
##   between, lag, lead
##
## Please cite as:
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables and Models
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
data("wagepan");help(wagepan)
head(wagepan[0:5])
```

```
##   nr year agric black bus
## 1 13 1980     0     0   1
## 2 13 1981     0     0   0
## 3 13 1982     0     0   1
## 4 13 1983     0     0   1
## 5 13 1984     0     0   0
## 6 13 1985     0     0   1
```

データは F. Vella and M. Verbeek (1998), “Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men,” で用いられているものを使う。労働組合の活動による賃金上昇の男女差?を研究した論文らしい。

R でのパネルデータ解析には plm パッケージを用いる。

まずは縦持ちになっている data.frame をパネルデータに変換する

```
p.df <- pdata.frame(wagepan, index=c("nr", "year"))
pdim(p.df) # dimensionを確認
```

```
## Balanced Panel: n = 545, T = 8, N = 4360
```

Fixed Effect Model の推定には、plm 関数で model="within" を指定すればよい

```
reg.fe <- plm(lwage~married+union+factor(year)*educ,
              data=p.df, model="within")
summary(reg.fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
```

```
## plm(formula = lwage ~ married + union + factor(year) * educ,
##      data = p.df, model = "within")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -4.152111 -0.125630  0.010897  0.160800  1.483401
##
## Coefficients:
##
##              Estimate Std. Error t-value Pr(>|t|)
## married          0.0548205  0.0184126  2.9773  0.002926 **
## union            0.0829785  0.0194461  4.2671  2.029e-
05 ***
## factor(year)1981    -0.0224158  0.1458885 -0.1537  0.877893
## factor(year)1982    -0.0057611  0.1458558 -0.0395  0.968495
## factor(year)1983     0.0104297  0.1458579  0.0715  0.942999
## factor(year)1984     0.0843743  0.1458518  0.5785  0.562965
## factor(year)1985     0.0497253  0.1458602  0.3409  0.733190
## factor(year)1986     0.0656064  0.1458917  0.4497  0.652958
## factor(year)1987     0.0904448  0.1458505  0.6201  0.535216
## factor(year)1981:educ 0.0115854  0.0122625  0.9448  0.344827
## factor(year)1982:educ 0.0147905  0.0122635  1.2061  0.227872
## factor(year)1983:educ 0.0171182  0.0122633  1.3959  0.162830
## factor(year)1984:educ 0.0165839  0.0122657  1.3521  0.176437
## factor(year)1985:educ 0.0237085  0.0122738  1.9316  0.053479 .
## factor(year)1986:educ 0.0274123  0.0122740  2.2334  0.025583 *
## factor(year)1987:educ 0.0304332  0.0122723  2.4798  0.013188 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    572.05
## Residual Sum of Squares: 474.35
## R-Squared:              0.1708
## Adj. R-Squared: 0.048567
## F-statistic: 48.9069 on 16 and 3799 DF, p-value: < 2.22e-16
```

## ランダム効果モデル

(板書)

### Rでの実装

pvar 関数で時間変動、個人変動の有無を変数ごとに確認できる

```
pvar(p.df)
```

```
## no time variation:      nr black hisp educ
## no individual variation: year d81 d82 d83 d84 d85 d86 d87
```

モデルの推定自体は固定効果モデルと同様に行うことができる。比較のため、OLSを含めてstargazerで出力する。

```
p.df$yr <- p.df$year %>% as.factor()
reg.ols <- plm(lwage~educ+black+hisp+exper+I(exper^2)+married+union+yr,
              data=p.df, model='pooling')
reg.re <- plm(lwage~educ+black+hisp+exper+I(exper^2)+married+union+yr,
              data=p.df, model='random')
reg.fe <- plm(lwage~I(exper^2)+married+union+yr,
              data=p.df, model="within")
stargazer(reg.ols, reg.re, reg.fe, type='text',
           column.labels=c("OLS", "RE", "FE"), keep.stat=c("n", "rsq"),
           keep=c("educ", "bl", "hi", "exp", "mar", "uni"))
```

```
##
## =====
##               Dependent variable:
##               -----
##               lwage
##               OLS      RE      FE
##               (1)      (2)      (3)
## -----
## educ          0.091***  0.092***
##               (0.005)  (0.011)
##
## black         -0.139*** -0.139***
##               (0.024)  (0.048)
##
## hisp          0.016     0.022
```

```
##          (0.021)  (0.043)
##
## exper      0.067***  0.106***
##          (0.014)  (0.015)
##
## I(exper2)  -0.002*** -0.005*** -0.005***
##          (0.001)  (0.001)  (0.001)
##
## married    0.108***  0.064***  0.047**
##          (0.016)  (0.017)  (0.018)
##
## union      0.182***  0.106***  0.080***
##          (0.017)  (0.018)  (0.019)
##
## -----
## Observations  4,360      4,360      4,360
## R2            0.189      0.181      0.181
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

また、モデルの特定化検定は、ハウスマン検定を用いることができる。

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

```
##
## Hausman Test
##
## data:  lwage ~ I(exper^2) + married + union + yr
## chisq = 26.361, df = 10, p-value = 0.003284
## alternative hypothesis: one model is inconsistent
```

有意水準5%以下でランダム効果モデルが正しい(個別効果の平均は0である)  
という帰無仮説が棄却された(板書)