Pitkittäis- ja paneeliaineistojen analysointi
Laskuharjoitus 7

2014, Huhtikuu 30

1 Laskutehtäviä

1. Tutkitaan lineaarista mallia

$$\log \text{wage}_{it} = \theta_1 + \theta_2 d2_t + Z_{it}\gamma + \delta_1 \text{female}_i + \delta_2 \cdot d2_t \cdot \text{female}_i + c_i + u_{it},$$

t=1,2, missä wage_{it} on henkilön i palkka vuonna t, $\theta_1, \theta_2 \in \mathbf{R}$, $d2_t = 0$, kun $t = 1$ ja $d2_t = 1$, kun $t = 2$, Z_{it} on $1 \times K$ vektori jossa on selittäviä muuttujia, γ on $K \times 1$-vektori, $\delta_1 \in \mathbf{R}$, $\text{female}_i = 1$, jos henkilö i on nainen ja $\text{female}_i = 0$, jos henkilö i on mies, $\delta_2 \in \mathbf{R}$, $c_i \in \mathbf{R}$ on tekijä jota ei havaita ja $u_{it} \in \mathbf{R}$ on virhetermi.

Millainen lineaarinen malli saadaan aikadifferenssille

$$\Delta \log \text{wage}_{i2} = \log \text{wage}_{i2} - \log \text{wage}_{i1}?$$

2. Tutkitaan lineaarista mallia

$$Y_{it} = \theta_1 + \theta_2 d2_t + \delta \text{prog}_{it} + c_i + u_{it},$$

t=1,2, missä $Y_{it} = \log \text{wage}_{it}$, $\theta_1, \theta_2 \in \mathbf{R}$, $d2_t = 0$, kun $t = 1$ ja $d2_t = 1$, kun $t = 2$, $\delta \in \mathbf{R}$, $\text{prog}_{it} = 1$, jos henkilö i osallistui koulutusohjelmaan vuonna t, $c_i \in \mathbf{R}$ on tekijä jota ei havaita ja $u_{it} \in \mathbf{R}$ on virhetermi.

Olkoon $\text{prog}_{i1} = 0$ kaikille henkilöille i (alussa kukaan ei osallistunut koulutusohjelmaan) ja $\text{prog}_{i2} = 0$, jos henkilö i kuuluu kontrolliryhmään (ei osallistunut koulutusohjelmaan hetkellä 2) sekä $\text{prog}_{i2} = 1$, jos henkilö i osallistui koulutusohjelmaan hetkellä 2.

Muodosta lineaarinen malli differenssille $\Delta Y_i = Y_{i2} - Y_{i1}$ ja osoita, että tässä lineaarisessa mallissa OLS-estimaattori (pienimmän neliösumman estimaattori) parametrille δ on

$$\hat{\delta} = \overline{\Delta Y}_{treat} - \overline{\Delta Y}_{control},$$

missä keskimääräinen vaikutus koulutusohjelmaan osallistuneessa ryhmässä on

$$\overline{\Delta Y}_{treat} = \frac{\sum_{i=1}^{N} \Delta Y_i \cdot \text{prog}_{i2}}{\sum_{i=1}^{N} \text{prog}_{i2}}$$

ja keskimääräinen vaikutus kontrolliryhmässä on

$$\overline{\Delta Y}_{control} = \frac{\sum_{i=1}^{N} \Delta Y_i \cdot (1 - \text{prog}_{i2})}{\sum_{i=1}^{N} (1 - \text{prog}_{i2})},$$

missä $\Delta Y_i = Y_{i2} - Y_{i1}$.

3. Tarkastellaan kiinteitten vaikutusten mallia (fixed effects model)

$$Y_i \quad = \quad X_i \beta + c_i j_T + u_i,$$

$i = 1, \ldots, N$, missä Y_i ja u_i ovat $T \times 1$-vektoreita, X_i on $T \times K$-matriisi, β on $K \times 1$-vektori, $c_i \in \mathbf{R}$ ja j_T on $T \times 1$-vektori, jonka elementit ovat ykkösiä. Olkoon

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} X_i' \hat{\Omega}_{RE}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \hat{\Omega}_{RE}^{-1} Y_i \right),$$

missä

$$\hat{\Omega}_{RE} = \hat{\sigma}_u^2 I_T + \hat{\sigma}_c^2 j_T j_T'.$$

(a) Osoita, että

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} \check{X}_i' \check{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \check{X}_i' \check{Y}_i \right),$$

missä

$$\check{X}_i = C_T X_i, \qquad \check{Y}_i = C_T Y_i,$$

$$C_T = I_T - \hat{\lambda} \frac{1}{T} j_T j_T', \qquad \hat{\lambda} = 1 - \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + T \hat{\sigma}_c^2}.$$

Voit olettaa tunnetuksi, että

$$\hat{\Omega}_{RE}^{-1/2} = \frac{1}{\hat{\sigma}_u} C_T.$$

(b) Osoita, että

$$C_T X_i = X_i - \hat{\lambda}\bar{X}_i j_T.$$

4. Perustele tehtävän 3 avulla seuraavat väitteet.

   (a) Kun $\hat{\sigma}_c^2/\hat{\sigma}_u^2$ on suuri, niin $\hat{\beta}_{RE} \approx \hat{\beta}_{FE}$.

   (b) Kun $\hat{\sigma}_c^2/\hat{\sigma}_u^2 \approx 0$, niin $\hat{\beta}_{RE} \approx \hat{\beta}_{POLS}$.

   (c) Suurilla T:n arvoilla $\hat{\beta}_{RE} \approx \hat{\beta}_{FE}$.

2 Tietokonetehtäviä

Tutkitaaan aineistoa cornwell.raw, jossa on muuttujat

```
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
county          int    %9.0g                  county identifier
year            byte   %9.0g                  81 to 87
crmrte          float  %9.0g                  crimes committed per person
prbarr          float  %9.0g                  'probability' of arrest
prbconv         float  %9.0g                  'probability' of conviction
prbpris         float  %9.0g                  'probability' of prison sentenc
avgsen          float  %9.0g                  avg. sentence, days
polpc           float  %9.0g                  police per capita
density         float  %9.0g                  people per sq. mile
taxpc           float  %9.0g                  tax revenue per capita
west            byte   %9.0g                  =1 if in western N.C.
central         byte   %9.0g                  =1 if in central N.C.
urban           byte   %9.0g                  =1 if in SMSA
pctmin80        float  %9.0g                  perc. minority, 1980
wcon            float  %9.0g                  weekly wage, construction
wtuc            float  %9.0g                  wkly wge, trns, util, commun
wtrd            float  %9.0g                  wkly wge, whlesle, retail trade
wfir            float  %9.0g                  wkly wge, fin, ins, real est
wser            float  %9.0g                  wkly wge, service industry
wmfg            float  %9.0g                  wkly wge, manufacturing
wfed            float  %9.0g                  wkly wge, fed employees
wsta            float  %9.0g                  wkly wge, state employees
wloc            float  %9.0g                  wkly wge, local gov emps
mix             float  %9.0g                  offense mix: face-to-face/other
```

```
pctymle          float   %9.0g                      percent young male
d82              byte    %9.0g                      =1 if year == 82
d83              byte    %9.0g                      =1 if year == 83
d84              byte    %9.0g                      =1 if year == 84
d85              byte    %9.0g                      =1 if year == 85
d86              byte    %9.0g                      =1 if year == 86
d87              byte    %9.0g                      =1 if year == 87
lcrmrte          float   %9.0g                      log(crmrte)
lprbarr          float   %9.0g                      log(prbarr)
lprbconv         float   %9.0g                      log(prbconv)
lprbpris         float   %9.0g                      log(prbpris)
lavgsen          float   %9.0g                      log(avgsen)
lpolpc           float   %9.0g                      log(polpc)
ldensity         float   %9.0g                      log(density)
ltaxpc           float   %9.0g                      log(taxpc)
lwcon            float   %9.0g                      log(wcon)
lwtuc            float   %9.0g                      log(wtuc)
lwtrd            float   %9.0g                      log(wtrd)
lwfir            float   %9.0g                      log(wfir)
lwser            float   %9.0g                      log(wser)
lwmfg            float   %9.0g                      log(wmfg)
lwfed            float   %9.0g                      log(wfed)
lwsta            float   %9.0g                      log(wsta)
lwloc            float   %9.0g                      log(wloc)
lmix             float   %9.0g                      log(mix)
lpctymle         float   %9.0g                      log(pctymle)
lpctmin          float   %9.0g                      log(pctmin)
clcrmrte         float   %9.0g                      lcrmrte - lcrmrte[_n-1]
clprbarr         float   %9.0g                      lprbarr - lprbarr[_n-1]
clprbcon         float   %9.0g                      lprbconv - lprbconv[_n-1]
clprbpri         float   %9.0g                      lprbpri - lprbpri[t-1]
clavgsen         float   %9.0g                      lavgsen - lavgsen[t-1]
clpolpc          float   %9.0g                      lpolpc - lpolpc[t-1]
cltaxpc          float   %9.0g                      ltaxpc - ltaxpc[t-1]
clmix            float   %9.0g                      lmix - lmix[t-1]
```

Lue data R:ään komennoilla

```
file<-"http://cc.oulu.fi/~jklemela/panel/cornwell.raw"
data<-read.table(file=file)
```

5a. Estimoi lineaarinen malli FE-estimaattorilla (fixed effects estimaattorilla) käyttäen kaikkia vuosia 81-87. Mallissa vastemuuttuja on log($crmrte$)

ja selittävät muuttujat ovat $\log(prbarr)$, $\log(prbconv)$, $\log(prbpris)$, $\log(avgsen)$ ja $\log(polpc)$.

Ohje: Tarvitaan havainnot

$$\ddot{Y}_{it} = Y_{it} - \bar{Y}_i, \qquad \ddot{X}_{it} = X_{it} - \bar{X}_i,$$

joihin sovelletaan OLS-estimaattoria:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{Y}_i \right).$$

5b. Estimoi lineaarinen malli OLS-estimaattorilla (leat squares estimaattorilla) käyttäen kaikkia vuosia 81-87. Mallissa vastemuuttuja on $\log(crmrte)$. Selittävät muuttujat ovat $\log(prbarr)$, $\log(prbconv)$, $\log(prbpris)$, $\log(avgsen)$, $\log(polpc)$ sekä indikaattorimuuttujat otantayksiköille (county). Indikaattorimuuttujat (dummy variables) ovat

$$d1_i, \ldots, dN_i,$$

missä

$$d1_i = \begin{cases} 1, \text{ kun } i = 1, \\ 0, \text{ kun } i \neq 1, \end{cases} \ldots, dN_i = \begin{cases} 1, \text{ kun } i = N, \\ 0, \text{ kun } i \neq N. \end{cases}$$

(Indikaattorimuuttujia tarvitaan vain $N - 1$ kappaletta.)

Tarkista, että muuttujien $\log(prbarr)$, $\log(prbconv)$, $\log(prbpris)$, $\log(avgsen)$ ja $\log(polpc)$ kertoimien estimaatit ovat samat kuin tehtävässä 5a.

3 Kertauskysymyksiä (eivät kuulu laskuharjoitukseen)

1. Tarkastellaan kiinteitten vaikutusten mallia (fixed effects model)

$$Y_{it} = X_{it}\beta + c_i + u_{it},$$

$t = 1, \ldots, T$, $i = 1, \ldots, N$, missä $Y_{it}, u_{it} \in \mathbf{R}$, X_{it} on $1 \times K$-vektori, β on $K \times 1$-vektori ja $c_i \in \mathbf{R}$. Määrittele kiinteitten vaikutusten muunnos (fixed effects transformation).

2. Tarkastellaan kiinteitten vaikutusten mallia (fixed effects model)

$$Y_i \;=\; X_i\beta + c_i j_T + u_i,$$

$i = 1, \ldots, N$, missä Y_i ja u_i ovat $T \times 1$-vektoreita, X_i on $T \times K$-matriisi, β on $K \times 1$-vektori, $c_i \in \mathbf{R}$ ja j_T on $T \times 1$-vektori, jonka elementit ovat ykkösiä. Olkoon

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{Y}_i\right),$$

missä $\ddot{X}_i$ on $T \times K$-matriisi, jonka rivit ovat

$$\ddot{X}_{it} = X_{it} - \bar{X}_i, \qquad \bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$$

ja $\ddot{Y}_i$ on $T \times 1$-vektori, jonka alkiot ovat

$$\ddot{Y}_{it} = Y_{it} - \bar{Y}_i, \qquad \bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}.$$

Oletetaan, että (Y_i, X_i), $i = 1, \ldots, N$, ovat i.i.d. Oletetaan, että

$$E(u_{it} \,|\, X_i, c_i) = 0,$$

kun $t = 1, \ldots, T$. Oletetaan, että $E\ddot{X}_i'\ddot{X}_i = \sum_{t=1}^{T} E\ddot{X}_{it}'\ddot{X}_{it}$ on kääntyvä. Osoita, että

$$N^{1/2}(\hat{\beta}_{FE} - \beta) \xrightarrow{d} N\left(0, A^{-1}BA^{-1}\right),$$

kun $N \to \infty$, missä $A = E\ddot{X}_i'\ddot{X}_i$ ja $B = E\ddot{X}_i' u_i u_i' \ddot{X}_i$.

3. Tarkastellaan kiinteitten vaikutusten mallia (fixed effects model)

$$Y_i \;=\; X_i\beta + c_i j_T + u_i,$$

$i = 1, \ldots, N$, missä Y_i ja u_i ovat $T \times 1$-vektoreita, X_i on $T \times K$-matriisi, β on $K \times 1$-vektori, $c_i \in \mathbf{R}$ ja j_T on $T \times 1$-vektori, jonka elementit ovat ykkösiä. Olkoon

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{N} \Delta X_i' \Delta X_i\right)^{-1} \left(\sum_{i=1}^{N} \Delta X_i' \Delta Y_i\right),$$

missä ΔX_i on $(T-1) \times K$-matriisi, jonka rivit ovat

$$\Delta X_{it} = X_{it} - X_{i,t-1}, \qquad t = 2, \ldots, T,$$

ja ΔY_i on $(T-1) \times 1$-vektori, jonka alkiot ovat

$$\Delta Y_{it} = Y_{it} - Y_{i,t-1}, \qquad t = 2, \ldots, T.$$

Oletetaan, että (Y_i, X_i), $i = 1, \ldots, N$, ovat i.i.d. Oletetaan, että

$$E(u_{it} \,|\, X_i, c_i) = 0,$$

kun $t = 1, \ldots, T$. Oletetaan, että $E(\Delta X_i' \Delta X_i) = \sum_{t=2}^{T} E(\Delta X_{it}' \Delta X_{it})$ on kääntyvä. Osoita, että

$$N^{1/2}(\hat{\beta}_{FD} - \beta) \xrightarrow{d} N\left(0, A^{-1} B A^{-1}\right),$$

kun $N \to \infty$, missä $A = E(\Delta X_i' \Delta X_i)$ ja $B = E(\Delta X_i' \Delta u_i \Delta u_i' \Delta X_i)$.