REGRESSIOANALYYSIN JATKOKURSSI, 5–10 OP (aine- ja syventävät opinnot). 15.9.–16.12.2011. Kirjallisuus: Russell Davidson ja James MacKinnon: Econometric Theory and Methods. Luennoi: yliopistonlehtori Pekka Pere.

Harjoitukset 3 (pe 30.9.)

1. Pohditaan lineaarista mallia

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

jossa $\mathbf{u} \sim \mathsf{IID}(\mathbf{0}_{n\times 1}, \sigma^2 \mathbf{I}_n)$, $\mathbf{X}$ on täysiasteinen $n \times k$ -matriisi ($k \geq 1$), jonka ensimmäinen sarake koostuu ykkösistä ja $\boldsymbol{\beta} = [\beta_1 \ldots \beta_k]'$ on $k \times 1$-vektori. Voit käyttää vastauksissasi alla oleviin kysymyksiin PNS-menetelmään liittyviä yleisiä tuloksia johtamatta niitä. Perustele kaikki vastauksesi.

Tehtävä on klassinen eli selittää lapsen pituutta ($\mathbf{y}$) molempien vanhempien pituuksilla ($\mathbf{x}_2$ koostuu isien pituuksista x_{t2} ja $\mathbf{x}_3$ äitien pituuksista x_{t3} senttimetreissä; $k = 3$). Perustele, mitkä seuraavista regressiosta ovat i) mahdollisia (selittäjät eivät ole lineaarisesti sidottuja eli $\boldsymbol{\beta}$ on estimoitavissa yksikäsitteisesti) ja ii) tuottavat samat sovitteet ja residuaalit kuin edellä kuvattu regressio? (Kaikissa regressioissa on myös vakio selittäjänä.) Lapsen pituutta selitetään

1. isän pituudella ja isän ja äidin pituuden erotuksella eli $\mathbf{x}_2$:lla ja $(\mathbf{x}_2 - \mathbf{x}_3)$:lla.
2. äidin pituudella ja äidin ja isän pituuden erotuksella eli $\mathbf{x}_3$:lla ja $(\mathbf{x}_3 - \mathbf{x}_2)$:lla.
3. isän ja äidin pituudella sekä isän ja äidin pituuden erotuksella eli $\mathbf{x}_2$:lla, $\mathbf{x}_3$:lla ja $(\mathbf{x}_2 - \mathbf{x}_3)$:lla.
4. isän pituudella kerrottuna vakiolla 2:lla ja äidin pituudella kerrottuna $0,5$:llä eli $2\mathbf{x}_2$:lla ja $0,5\mathbf{x}_3$:lla.
5. isän ja äidin pituuden summalla $\mathbf{x}_2 + \mathbf{x}_3$.
6. isän ja äidin pituuden keskiarvolla $0,5(\mathbf{x}_2 + \mathbf{x}_3)$:lla.
7. isän pituudella ja isän ja äidin pituuden keskiarvolla eli $\mathbf{x}_2$:lla ja $0,5(\mathbf{x}_2 + \mathbf{x}_3)$:lla.
8. isän ja äidin pituudella sekä niiden tulolla eli $\mathbf{x}_2$:lla, $\mathbf{x}_3$:lla ja vektorilla, jonka t. elementti on $x_{t2}x_{t3}$.
9. isän ja äidin pituudella neliöitynä eli vektoreilla, joiden t:nnet elementit ovat x_{t2}^2 ja x_{t3}^2.
10. isän pituudella senttimetreissä ja äidin pituudella tuumissa eli $\mathbf{x}_2$:lla ja $0,3937\mathbf{x}_3$:lla.

Mikäli mahdollista määrittele selittäjien muunnokseen liittyvä yhtälöt $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\mathbf{A}\mathbf{A}^{-1}\boldsymbol{\beta}$ toteuttava $\mathbf{A}$-matriisi (3×3; kirjan s. 61).

2. Tutkitaan lineaarista mallia

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

jossa $\mathbf{u} \sim \mathsf{IID}(\mathbf{0}_{n\times 1}, \sigma^2 \mathbf{I}_n)$, $\mathbf{X}$ on täysiasteinen $n \times k$ -matriisi ($k \geq 1$), jonka ensimmäinen sarake koostuu ykkösistä ja $\boldsymbol{\beta} = [\beta_1 \ldots \beta_k]'$ on k-vektori. Voit käyttää vastauksissasi alla oleviin kysymyksiin PNS-menetelmään liittyviä yleisiä tuloksia johtamatta niitä. Perustele kaikki vastauksesi.

Lisätään selittäjiin (x_{it}) vakio $a_i \neq 0$ ($i = 2, \ldots, k$; $t = 1, \ldots, n$) kertomalla matriisi $\mathbf{X}$ oikealta matriisilla

$$\mathbf{A} = \begin{bmatrix} 1 & \mathbf{a}' \\ \mathbf{0}_{(k-1)\times 1} & \mathbf{I}_{k-1} \end{bmatrix},$$

jossa $\mathbf{a}' = [a_2 \ldots a_k]$. Uusi malli on

$$\mathbf{y} = \mathbf{X}\mathbf{A}\boldsymbol{\beta}^* + \mathbf{u}^* \tag{2}$$

ilmeisin merkinnöin.

Aputulos (perustele):

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & -\mathbf{a}' \\ \mathbf{0}_{(k-1)\times 1} & \mathbf{I}_{k-1} \end{bmatrix}.$$

a) Miten $\hat{\beta}_i^*$-estimaatit riippuvat estimaateista $\hat{\beta}_i (i = 2, \ldots, k)$? Esitä yksityiskohtainen kaava siitä, miten vakion estimaatti $\hat{\beta}_1^*$ riippuu estimaateista $\hat{\beta}_i$.

b) Eroavatko $\mathsf{S}(\mathbf{X})$ ja $\mathsf{S}(\mathbf{XA})$ eli selittäjämatriiseihin $\mathbf{X}$ ja $\mathbf{XA}$ liittyvät projektiomatriisit $\mathbf{P}_{\mathbf{X}}$ ja $\mathbf{P}_{\mathbf{XA}}$ toisistaan? Eroavatko estimoidut jäännökset toisistaan malleissa (1) ja (2)? Entä jäännösvarianssin estimaatit?

c) Poikkeavatko malleille (1) ja (2) lasketut selitysasteet R_c^2:t?

d) Palataan malliin (1). Oletetaan, että havainnot $\mathbf{y}$ kerrotaan vakiolla $c \neq 0$. Uusi malli on

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{u}^*, \tag{3}$$

jossa $\mathbf{y}^* = c\mathbf{y}$ ja muut merkinnät ovat ilmeiset. Miten $\hat{\beta}_i^*$-estimaatit riippuvat estimaateista $\hat{\beta}_i (i = 1, \ldots, k)$? Poikkeavatko malleille (1) ja (3) lasketut selitysasteet R_c^2:t?

3. Tehtävän merkinnät ovat kirjassa käytettyjä.

a) Osoita, että R_c^2 eli

$$\frac{\|\mathbf{P}_\mathbf{X}\mathbf{M}_\iota\mathbf{y}\|^2}{\|\mathbf{M}_\iota\mathbf{y}\|^2} = 1 - \frac{\|\mathbf{M}_\mathbf{X}\mathbf{y}\|^2}{\|\mathbf{M}_\iota\mathbf{y}\|^2},$$

kun vakiovektori $\boldsymbol{\iota} = [1 \ldots 1]'$ kuuluu $S(\mathbf{X})$:ään.

b) Voiko $\boldsymbol{\iota}$ kuulua $S(\mathbf{X})$:ään, jos mallissa ei ole vakiota?

4. Verrataan regressiota

$$\mathbf{y} = \boldsymbol{\iota}\boldsymbol{\beta}_0 + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_1^{'} \tag{1'}$$

ja keskistettyä regressiota

$$\mathbf{M}_\iota\mathbf{y} = \mathbf{M}_\iota\mathbf{X}\boldsymbol{\beta} + \mathbf{u}_1^*. \tag{1*}$$

Yllä $\boldsymbol{\iota} = [1 \ldots 1]'$, $\mathbf{X}$ on täysiasteinen $n \times k$ -matriisi, $\mathsf{S}(\mathbf{X})$ ei sisällä vakiovektoria, $\boldsymbol{\beta}$ on $k \times 1$ -vektori ja $\mathbf{u}_1^{'}$ ja $\mathbf{u}_1^*$ ovat tarkemmin määrittelemättömiä jäännösvektoreita.

a) Tulkitse regressiot sanoin. Tulkitse $\mathsf{S}(\mathbf{X})$:ää koskeva ehto sanoin molempien regressioiden tilanteessa.

b) Osoita, että $\boldsymbol{\beta}$:n PNS-estimaatti on sama molemmissa regressioissa. (Vihje: FWL-lause.)

c) Osoita, että estimoidut jäännökset $\hat{\mathbf{u}}_1^{'}$ ja $\hat{\mathbf{u}}_1^*$ ovat samat molemmissa regressioissa. (Vihje: FWL-lause.)

d) Osoita, että $\mathsf{R}_\mathsf{c}^2 = 1 - \|\mathbf{M}_\mathbf{X}\mathbf{y}\|^2 / \|\mathbf{M}_\iota\mathbf{y}\|^2$ on sama molemmissa regressioissa.

5.

a) Selitä kaavoin ja sanoin Frisch–Waugh–Lovell-lause.

b) Oheinen kuvio 1.7 havainnollistaa lausetta. Tulkitse regressioanalyysin käsittein kuvion 1.7 (a) vektorit *OA*, *OB*, *BA*, *OC*, *CA* ja *CB*.

c) Tulkitse regressioanalyysin käsittein kuvion 1.7 (b) vektorit *OB*, *OE*, *OD*, *OC* ja *CB*.

d) Selitä sanoin, miten kuvion 1.7 (c) vektorit liittyvät Frisch–Waugh–Lovell-lauseeseen.