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 7 (pe 4.11.)

1. Johdetaan aputuloksia seuraavaa tehtävää varten. Tutkitaan keskistettyä regressiota

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

Yllä $\boldsymbol{\iota} = [1 \dots 1]'$, $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$ on täysiasteinen $n \times k$ -matriisi, $\mathbf{X}_1$ ja $\mathbf{X}_2$ ovat $n \times k_1$ ja $n \times k_2$ -matriiseja ($k_1 + k_2 = k$), $\mathsf{S}(\mathbf{X})$ ei sisällä vakiovektoria, $\boldsymbol{\beta}$ on $k \times 1$ -vektori ja $\mathbf{u}_1^*$ on tarkemmin määrittelemätön residuaalivektori.

a) Todista, että regressiossa (1*) selitysaste R_c^2 voidaan laskea kaavalla

$$\mathsf{R}_\mathsf{c}^2 \equiv 1 - \frac{\|\mathbf{M}_{\mathbf{X}^*}\mathbf{y}^*\|^2}{\|\mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}^*\|^2} = \frac{\|\mathbf{P}_{\mathbf{X}^*}\mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}^*\|^2}{\|\mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}^*\|^2},$$

jossa $\mathbf{y}^* = \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}$ ja $\mathbf{X}^* = \mathbf{M}_{\boldsymbol{\iota}}\mathbf{X}$. Todista siis, että yhtäsuuruus pätee yllä, kun regression muuttujat ovat keskistettyjä.

b) Osoita, että pätee

$$\mathbf{P}_2\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1} = \mathbf{0}_{n\times n},$$

c) Osoita, että kun $\mathbf{y}$ on keskistetty, pätee

$$\mathbf{M}_{\boldsymbol{\iota}}\mathbf{y} = \mathbf{y}.$$

d) Olkoon $\mathbf{y}'\mathbf{M}_2\mathbf{X}_1 = \mathbf{0}_{1\times k_1}$. Osoita, että tällöin

$$\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1}\mathbf{y} = \mathbf{0}_{n\times 1}.$$

e) Olkoon $\mathbf{y}'\mathbf{X}_1 = \mathbf{0}_{1\times k1}$. Osoita, että tällöin

$$\mathbf{P}_1\mathbf{y} = \mathbf{0}_{n\times 1}.$$

ja

$$\mathsf{R}_\mathsf{c}^2(\mathbf{X}_1) = 0.$$

Yllä $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_1)$ on R_c^2 laskettuna regressiosta (2) alla, kun $\mathbf{y}$ ja $\mathbf{X}_1$ ovat keskistettyjä.

f) Olkoon $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}_{k_1\times k_2}$. Osoita, että tällöin

$$\mathbf{M}_1\mathbf{X}_2 = \mathbf{X}_2,$$

$$\mathbf{P}_1\mathbf{P}_2 = \mathbf{0}_{n\times n}$$

ja

$$\mathbf{P}_{\mathbf{M}_1\mathbf{X}_2} = \mathbf{P}_2.$$

2. Nyt itse asiaan! Tutkitaan regressioita

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}_1, \tag{1}$$

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u}_2 \tag{2}$$

ja

$$\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}_3. \tag{3}$$

Pohditaan, miten selitysaste R_c^2 voi käyttäytyä regressioissa (1), (2) ja (3). Merkitään selitysasteita niissä $\mathsf{R}_\mathsf{c}^2(\mathbf{X})$:llä, $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_1)$:llä ja $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_2)$:lla. Oletetaan lisäksi, että $\mathsf{R}_\mathsf{c}^2(\mathbf{X})$ on suuri, $\mathsf{S}(\mathbf{X}_1, \mathbf{X}_2)$ ei sisällä vakiovektoria ja että muuttujat ovat keskistettyjä kuten regressiossa (1*), mutta kaavojen yksinkertaistamiseksi ei osoiteta keskistämistä ylimääräisillä merkinnöillä. Keskistäminen helpottaa tehtävään liittyviä laskutoimituksia.

Väitteet: On mahdollista, että

$$\mathsf{R}_\mathsf{c}^2(\mathbf{X}) - \mathsf{R}_\mathsf{c}^2(\mathbf{X}_1) < \mathsf{R}_\mathsf{c}^2(\mathbf{X}_2), \tag{a}$$

$$\mathsf{R}_\mathsf{c}^2(\mathbf{X}) - \mathsf{R}_\mathsf{c}^2(\mathbf{X}_1) > \mathsf{R}_\mathsf{c}^2(\mathbf{X}_2) \tag{b}$$

tai

$$\mathsf{R}_\mathsf{c}^2(\mathbf{X}) - \mathsf{R}_\mathsf{c}^2(\mathbf{X}_1) = \mathsf{R}_\mathsf{c}^2(\mathbf{X}_2). \tag{c}$$

On siis mahdollista, että

a) sekä $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_1)$ että $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_2)$ ovat suuria. Tällöin $\mathbf{X}_1$ vaikuttaa yksin tärkeältä selittäjämatriisilta, mutta sen poistaminen regressiosta (1) pienentää vain vähän selitysastetta!

b) sekä $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_1)$ että $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_2)$ ovat pieniä vaikka $\mathsf{R}_\mathsf{c}^2(\mathbf{X})$ on suuri. Tällöin $\mathbf{X}_1$ vaikuttaa yksin epäoleelliselta selittäjämatriisilta, mutta sen poistaminen regressiosta (1) romahduttaa selitysasteen!

c) $\mathsf{R}_\mathsf{c}^2(\mathbf{X})$ on $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_1)$:n ja $\mathsf{R}_\mathsf{c}^2(\mathbf{X}_2)$:n summa.

Todista analyyttisesti väitteet a)–c). Vihje1: HT:t 5.2 ja 7.1. Vihje2: Oleta

a) $\mathbf{y}'\mathbf{M}_2\mathbf{X}_1 = \mathbf{0}_{1\times k_1}$ (ja $\mathbf{y}'\mathbf{X}_1 \neq \mathbf{0}_{1\times k_1}$).

b) $\mathbf{y}'\mathbf{X}_1 = \mathbf{0}_{1\times k_1}$ (ja $(\mathbf{M}_2\mathbf{X}_1)'\mathbf{y} \neq \mathbf{0}_{n\times 1}$).

c) $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}_{k_1\times k_2}$ (ja $\mathbf{y}'\mathbf{X}_1 \neq \mathbf{0}_{1\times k_1}$ ja $\mathbf{y}'\mathbf{X}_2 \neq \mathbf{0}_{1\times k_2}$).

3. Tutkitaan lineaarista mallia

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{4}$$

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 $\boldsymbol{\beta}$ on $k \times 1$ -vektori. Lasketaan uusiksi selittäjiksi lineaarikombinaatioita alkuperäisistä kertomalla matriisi $\mathbf{X}$ oikealta ei-singulaarisella matriisilla $\mathbf{A}$ ($k \times k$). Uusi malli on

$$\mathbf{y} = \mathbf{X}\mathbf{A}\boldsymbol{\beta}^* + \mathbf{u}^* \tag{5}$$

ilmeisin merkinnöin.

a) Mikä on PNS-estimaatti $\boldsymbol{\beta}$:lle mallista (1) ja $\boldsymbol{\beta}^*$:lle mallista (5)? Osoita, että $\hat{\boldsymbol{\beta}}^* = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$.

b) Eroavatko $\mathsf{S}(\mathbf{X})$ ja $\mathsf{S}(\mathbf{XA})$ eli selittäjämatriiseihin $\mathbf{X}$ ja $\mathbf{XA}$ liittyvät projektiomatriisit $\mathbf{P_X}$ ja $\mathbf{P_{XA}}$ toisistaan? Eroavatko estimoidut jäännökset toisistaan malleissa (4) ja (5)? Entä jäännösvarianssin estimaatit?

c) Poikkeavatko malleille (4) ja (5) lasketut selitysasteet R^2_c:t? Perustele.

d) Laske $\hat{\boldsymbol{\beta}}$:n ja $\hat{\boldsymbol{\beta}}^*$:n kovarianssimatriisit.

e) Osoita, että $\hat{\boldsymbol{\beta}}^*$:n yksittäisen elementin t-arvo ($t_{\beta_i^*=0} = \hat{\beta}_i^* / \mathsf{SE}(\hat{\beta}_i^*)$) on $\mathbf{A}_i^{-1}\hat{\boldsymbol{\beta}}/s\{\mathbf{A}_i^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}_i^{-1})'\}^{1/2}$, jossa $\mathbf{A}_i^{-1}$ on $\mathbf{A}^{-1}$:n i. vaakarivi ja s on estimoitu jäännösten keskihajonta. (Vihje: Matriisin i. diagonaalitermi saadaan kertomalla matriisi vasemmalta ja oikealta vektoreilla $\mathbf{e}_i' = [0\ldots0 \quad 1 \quad 0\ldots0]$ ja $\mathbf{e}_i$. Edellä "1" on i. termi.)

f) Jos $\mathbf{A}$ on diagonaalimatriisi $\lceil a_1 \ldots a_k \rfloor$ eli alkuperäiset muuttujat vain skaalataan, niin miten estimaatit yhtälöistä (4) ja (5) riippuvat toisistaan? Muuttuvatko estimoitujen kertoimien t-arvot? Perustele.

4.

a) Olkoon regressiomalli

$$y_t = \beta t^{-1} + u_t, \quad u_t \sim \mathsf{IID}(0, \sigma^2),$$

jossa β^0 on parametrin β todellinen arvo ja $t = 1, \ldots, n$ (oleellisesti kirjan malli (3.20)). Perustele, että β:n PNS-estimaattori ei ole tarkentuva. Selitä tulos intuitiivisesti. (Vihje1: $\lim_{n\to\infty} \sum_{t=1}^{n} t^{-2} = \pi^2/6$. Vihje 2: Osoita, että PNS-estimaattorin varianssi ei ole nolla edes asymptoottisesti eli kun $n \to \infty$.)

b) Olkoon regressiomalli

$$y_t = \beta_1 + \beta_2 t^{-1} + u_t, \quad u_t \sim \mathsf{IID}(0, \sigma^2),$$

jossa β_1^0 ja β_2^0 ovat parametrien β_1 ja β_2 todelliset arvot ja $t = 1, \ldots, n$ (kirjan malli (3.20)). Todista, että tehtävän tilanteessa parametrivektori $\boldsymbol{\beta} = [\beta_1 \ \beta_2]'$ ei ole asymptoottisesti identifioituva PNS-estimoinnissa eli että

$$\boldsymbol{\alpha}(\boldsymbol{\beta}) = \mathsf{plim}\, n^{-1}\mathbf{W}[\mathbf{y} - \mathbf{x}(\boldsymbol{\beta})] = \begin{bmatrix} \beta_1^0 - \beta_1 \\ 0 \end{bmatrix} \neq \mathbf{0},$$

jossa $\mathbf{W}_t = [1 \ t^{-1}]$ ja $x_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 t^{-1}$. (Kirjan merkinnät.) (Vihje1: $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} t^{-1} = 0$, $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} t^{-2} = 0$, $\mathsf{plim} n^{-1} \sum_{t=1}^{n} u_t = 0$ ja $\mathsf{plim} n^{-1} \sum_{t=1}^{n} u_t/t = 0$.) Onko PNS-estimaattori $\boldsymbol{\beta}$:lle siten tarkentuva?