REGRESSIOANALYYSIN TEORIAN JATKOKURSSI, 5 OP (syventävät opinnot). 4.9.–10.10.2008. Kirjallisuus: Russell Davidson ja James MacKinnon: Econometric Theory and Methods, luvut 1–7. Luennoi: yliopistonlehtori Pekka Pere.

Uusintakoe yleistentissä 12.11.2008

Vastaa kolmeen kysymyksistä 1–4. Jos vastaat useampaan, merkitse selkeästi, mitkä kolme kysymystä tahdot arvosteltavan. Tehtävät ovat kuuden pisteen arvoisia. Palauta kysymyspaperi.

1. 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 $\boldsymbol{\beta} = [\beta_1 \dots \beta_k]'$ on $k \times 1$-vektori. Voit käyttää vastauksissasi alla oleviin kysymyksiin PNS-menetelmään liittyviä yleisiä tuloksia johtamatta niitä. Perustele kaikki vastauksesi.

Lisätään selittäjiin (x_{it}) vakio $a_i \neq 0$ ($i = 2, \dots, k$; $t = 1, \dots, n$) 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 $\mathbf{a}' = [a_2 \dots a_k]$. Uusi malli on

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

ilmeisin merkinnöin.

a) Miten $\hat{\beta}_i^*$-estimaatit riippuvat estimaateista $\hat{\beta}_i (i = 2, \dots, k)$? 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 $\mathbf{P}_{\mathbf{X}^*}$ toisistaan? Eroavatko estimoidut jäännökset toisistaan malleissa (1) ja (2)? Entä jäännösvarianssin estimaatit?

c) Poikkeavatko malleille (1) ja (2) lasketut selitysasteet R_c^2:t? Millaisessa suhteessa ovat F-testisuureet nollahypoteeseille H_0: $\beta_2 = \dots \beta_k = 0$ ja H_0^*: $\beta_2^* = \dots \beta_k^* = 0$?

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? Millaisessa suhteessa ovat F-testisuureet nollahypoteeseille H_0: $\beta_2 = \ldots \beta_k = 0$ ja H_0^*: $\beta_2^* = \ldots \beta_k^* = 0$?

2.

Tutkitaan lineaarista mallia

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}, \tag{4}$$

jossa $\mathbf{u} \sim \mathsf{IID}(\mathbf{0}_{n\times 1}, \sigma^2\mathbf{I}_n)$, $\mathbf{X}_1$ ja $\mathbf{X}_2$ ovat täysiasteisia $n \times k_1$ ja $n \times k_2$ -matriiseja ja $\boldsymbol{\beta}_1$ on $k_1 \times 1$ ja $\boldsymbol{\beta}_2$ on $k_2 \times 1$ -vektori.

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

b) Olkoon DGP mallin (4) mukainen mutta estimoidaan rajoitettu malli

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u}^* \tag{5}$$

ilmeisin merkinnöin.

Olkoon $\boldsymbol{\beta}_2 = \mathbf{0}$. Perustele, että $\boldsymbol{\beta}_1$:n PNS-estimaattori $\tilde{\boldsymbol{\beta}}_1$ mallista (5) estimoituna on tehokkaampi (tai ainakin yhtä tehokas) kuin $\hat{\boldsymbol{\beta}}_1$ eli PNS-estimaattori $\boldsymbol{\beta}_1$:lle mallista (4) estimoituna.

Olkoon $\boldsymbol{\beta}_2 \neq \mathbf{0}$. Osoita, että $\boldsymbol{\beta}_1$:n PNS-estimaattori $\tilde{\boldsymbol{\beta}}_1$ mallista (5) estimoituna ei ole yleensä harhaton estimaattori $\boldsymbol{\beta}_1$:lle. Milloin $\tilde{\boldsymbol{\beta}}_1$ on harhaton estimaattori $\boldsymbol{\beta}_1$:lle? Voiko $\tilde{\boldsymbol{\beta}}_1$ olla keskineliövirheen mielessä tarkempi estimaattori kuin $\hat{\boldsymbol{\beta}}_1$? Perustele.

3. Tutkitaan lineaarista regressiomallia

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

jossa $\mathsf{E}(\mathbf{u}) = \mathbf{0}$ ja $\mathsf{E}(\mathbf{u}\mathbf{u}') = \lceil \omega_1^2 \ldots \omega_n^2 \rfloor$ (diagonaalimatriisi) ja muut merkinnät ovat ilmeiset.

a) Selitä sanoin ja kaavoin, mikä on Whiten heteroskedastisuus-tarkentuva kovarianssimatriisi -estimaattori. Miten se liittyy malliin yllä?

b) Selitä sanoin ja kaavoin Gauss–Newton-regressio (GNR) ja miten se liittyy Gauss-Newton -menetelmään.

c) Miten mallin jäännöksen heteroskedastisuutta voidaan testata Gauss–Newton-regression avulla?

4. Tutkitaan mallia

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad \mathbf{u} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Omega}), \qquad \boldsymbol{\Omega}^{-1} = \boldsymbol{\Psi}\boldsymbol{\Psi}'$$

(kirjan merkinnöillä).

a) Selitä merkinnät yllä huolellisesti. Johda yleistetty PNS-estimaattori (GLS estimator) parametrivektorille $\boldsymbol{\beta}$. Perustele kaikki vaiheet huolellisesti.

b) Olkoon $\boldsymbol{\Omega}$ diagonaalimatriisi. Mihin muotoon yleistetty PNS-estimaattori yksinkertaistuu? Selitä intuitiivisesti, miksi yleistetty PNS-estimaattori on tarkempi kuin PNS-estimaattori.

c) Mitä tarkoitetaan toteutettavissa olevalla (feasible) yleistetyllä PNS-estimaattorilla? Millaisia ominaisuuksia (ongelmia?) tällaisella estimaattorilla on? Selitä sen laskeminen vaihe vaiheelta perustellen erityistilanteessa, jossa $\boldsymbol{\Omega}$ on diagonaalimatriisi ja

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \qquad \mathsf{E}(u_t^2) = \exp(\mathbf{Z}_t\gamma)$$

(kirjan merkinnöillä).

APUTULOKSIA

Käänteismatriiseja koskevia laskusääntöjä:

- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.
- 1. tehtävän tilanteessa

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

Definittisyyssääntöjä:

- $\mathbf{CC}'$ on positiivisesti semidefiniitti matriisi.
- $(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}$ (kirjan merkinnöillä) on positiivisesti semidefiniitti matriisi.

Muuta: Kun β_1 on vakion kerroin, pätee

$$F_{\beta_2=0} = \frac{n-k}{k-1} \times \frac{\mathsf{R}_c^2}{1-\mathsf{R}_c^2}.$$

REGRESSIOANALYYSIN TEORIAN JATKOKURSSI, 5 OP (syventävät opinnot). 4.9.–10.10.2008. Kirjallisuus: Russell Davidson ja James MacKinnon: Econometric Theory and Methods, luvut 1–7. Luennoi: yliopistonlehtori Pekka Pere.

Yleistenttioppukoe 3.3.2009

Vastaa kolmeen kysymyksistä 1–4. Jos vastaat useampaan, niin merkitse selkeästi, mitkä kolme kysymystä tahdot arvosteltavan. Tehtävät ovat kuuden pisteen arvoisia. Palauta kysymyspaperi.

1. 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 $\boldsymbol{\beta}$ 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

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

ilmeisin merkinnöin. Voit käyttää vastauksissasi alla oleviin kysymyksiin PNS-menetelmään liittyviä yleisiä tuloksia johtamatta niitä.

a) Mikä on PNS-estimaatti $\boldsymbol{\beta}$:lle mallista (1) ja $\boldsymbol{\beta}^*$:lle mallista (2)? 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 (1) ja (2)? Entä jäännösvarianssin estimaatit?

c) Poikkeavatko malleille (1) ja (2) lasketut selitysasteet R_c^2: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 (1) ja (2) riippuvat toisistaan? Muuttuvatko estimoitujen kertoimien t-arvot? Perustele.

2.

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

3. Tutkitaan lineaarista regressiomallia

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

jossa $\mathsf{E}(\mathbf{u}) = \mathbf{0}$ ja $\mathsf{E}(\mathbf{u}\mathbf{u}') = \lceil \omega_1^2 \ldots \omega_n^2 \rfloor$ (diagonaalimatriisi) ja muut merkinnät ovat ilmeiset.

a) Selitä sanoin ja kaavoin, mikä on Whiten heteroskedastisuus-tarkentuva kovarianssimatriisi -estimaattori. Miten se liittyy malliin yllä?

b) Selitä sanoin ja kaavoin Gauss–Newton-regressio (GNR) ja miten se liittyy Gauss-Newton -menetelmään.

c) Miten mallin jäännöksen heteroskedastisuutta voidaan testata Gauss–Newton-regression avulla?

4. Tutkitaan mallia

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad \mathbf{u} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Omega}), \qquad \boldsymbol{\Omega}^{-1} = \boldsymbol{\Psi}\boldsymbol{\Psi}'$$

(kirjan merkinnöillä).

a) Selitä merkinnät yllä huolellisesti. Johda yleistetty PNS-estimaattori (GLS estimator) parametrivektorille $\boldsymbol{\beta}$. Perustele kaikki vaiheet huolellisesti.

b) Olkoon $\boldsymbol{\Omega}$ diagonaalimatriisi. Mihin muotoon yleistetty PNS-estimaattori yksinkertaistuu? Selitä intuitiivisesti, miksi yleistetty PNS-estimaattori on tarkempi kuin PNS-estimaattori.

c) Mitä tarkoitetaan toteutettavissa olevalla (feasible) yleistetyllä PNS-estimaattorilla? Millaisia ominaisuuksia (ongelmia?) tällaisella estimaattorilla on? Selitä sen laskeminen vaihe vaiheelta perustellen erityistilanteessa, jossa $\boldsymbol{\Omega}$ on diagonaalimatriisi ja

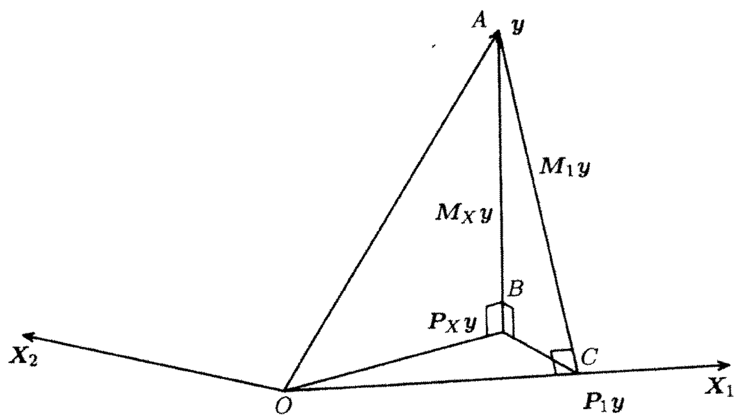
$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \qquad \mathsf{E}(u_t^2) = \exp(\mathbf{Z}_t\gamma)$$

(kirjan merkinnöillä).

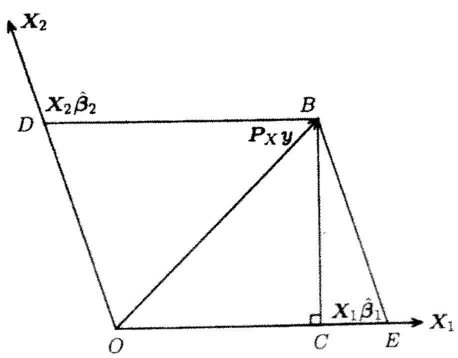
APUTULOKSIA

Käänteismatriiseja koskevia laskusääntöjä:

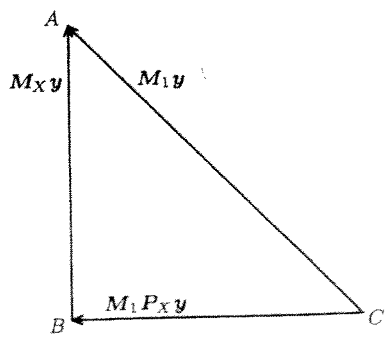
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.



(a) Two projections of $\boldsymbol{y}$



(b) The span $\mathcal{S}(\boldsymbol{X})$ of the regressors



(c) The decomposition of $\boldsymbol{M}_1\boldsymbol{y}$

Figure 1.7 The Frisch-Waugh-Lovell Theorem