Ekonometrinė panelinių duomenų analizė

Vaidotas Zemlys

Matematikos ir Informatikos Fakultetas, Ekonometrinės analizės katedra
Vaidotas.Zemlys@maf.vu.lt

2005 m. kovo 15 d.

Paneliniai Duomenys

- Duomenys apie objektus (namų ūkius, firmas, šalis) surinkti keliems laiko periodams.
- Panel Study of Income Dynamics (PSID) - duomenys apie US namų ūkius. Pradėti rinkti nuo 1968 metų apie 4802 šeimas, apklausiant jas kiekvienais metais.
- Nuo 80-tųjų panašios apklausos buvo pradėtos atlikti kitose šalyse: Vokietijoje, Nyderlanduose, Belgijoje, Kanadoje, Prancūzijoje, Anglijoje.

Panelinių duomenų privalumai

- Atsižvelgiama į individualų heterogeniškumą.
- Daugiau informacijos, kintamumo, mažiau kolinearumo tarp kintamųjų, daugiau laisvės laipsnių ir efektyvumo.
- Labiau pritaikyti studijuoti prisitaikymo dinamiką.
- Galima nustatyti ir išmatuoti dydžius, kurių neįmanoma pastebėti turint grynai pjūvinius (cross-section) arba grynai laiko eilučių duomenis.
- Įmanoma kontstruoti ir testuoti labiau komplikuotus elgsenos modelius, nei turint pjūvinius arba laiko eilučių duomenis.
- Duomenys renkami mikro lygyje, todėl išvengiama paklaidų dėl agregavimo.

Panelinių duomenų trūkumai

- Plano ir duomenų surinkimo problemos.
- Matavimo paklaidos turi didesnę įtaką.
- Pasirenkamumo problemos
- Trumpa laiko eilučių dimensija.

Vieno paklaidos komponento regresijos modelis

- Regresijos modelis paneliniams duomenims:

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \quad i = 1, \dots, N; \; t = 1, \dots, T$$

- Dažniausiai naudojamas vieno paklaidos komponento modelis

$$u_{it} = \mu_i + \nu_{it},$$

  čia μ_i nestebimas indvidualus poveikis.

- Matricinis modelio pavidalas

$$y = \alpha\imath_{NT} + X\beta + Z_\mu\mu + \nu,$$

  čia $Z_\mu = I_N \otimes \imath_T$.

Vidinė ir tarpinė regresija

- Vidinės regresijos β įvertis gaunamas iš regresijos

$$Qy = QX\beta + Q\nu,$$

čia $Q = I_{NT} - I_N \otimes \bar{J}_T$ - simetrinė idempotentinė matrica. Tipiškas Qy elementas yra $y_{it} - \bar{y}_{i.}$.

- Tarpinės regresijos įvertis gaunamas iš regresijos

$$Py = \alpha + PX\beta + P\nu,$$

čia $P = I_N \otimes \bar{J}_T$ - simetrinė idempotentinė matrica. Tipiškas Py elementas yra $\bar{y}_{i.}$, pakartotas T kartų kiekvienam individui.

- $P + Q = I$, $PQ = 0$.

Fiksuotų efektų modelis

- μ_i - fiksuoti. Įverčius gauname taikydami LSDV.
- LSDV įverčių gavimą iliustruoja šios lygtys, atveju, kai turime tik vieną nepriklausomą kintamąjį:

$$\begin{aligned} \bar{y}_{i.} &= \alpha + \beta\bar{x}_{i.} + \mu_i + \bar{\nu}_{i.} \\ y_{it} - \bar{y}_{i.} &= \beta(x_{it} - \bar{x}_{i.}) + (\nu_{it} - \bar{\nu}_{i.}) \\ \bar{y}_{..} &= \alpha + \beta\bar{x}_{..} + \bar{\nu}_{..} \end{aligned}$$

- Imamas vidinės regresijos įvertis $\tilde{\beta}$, gaunamas iš antros lygties. α įvertinamas iš trečios lygties, taikant sąlygą $\sum \mu_i = 0$. μ_i įverčiai gaunami iš pirmos lygties. Jei $\sum \mu_i \neq 0$, galėsime įvertinti tik $\alpha + \mu_i$

FE modelio apžvalga

- Tinka, jei nagrinėjame tam tikrą specifinę objektų grupę (pvz. N firmų, N šalių) ir ja remdamiesi norime daryti tam tikras išvadas.
- Jei tikrasis modelis yra FE, LSDV yra BLUE, jei $\nu_{it} \sim IID(0, \sigma_\nu^2)$.
- Jei $T \to \infty$, tai $LSDV$ suderintas, jei T – fiksuotas, o $N \to \infty$, tai suderintas tik β įvertis.
- Jei tikrasis modelis FE, tai taikant įprastą MKM gausime paslinktus įverčius.
- Jei tarp nepriklausomų kintamųjų yra tokių, kurie nekinta laike, tai jų įtakos su LSDV neįvertinsime.

Atsitiktinių efektų modelis

- Prielaidos: $\mu_i \sim IID(0, \sigma_\mu^2)$, $\nu_{it} \sim IID(0, \sigma_\nu^2)$
- Paklaidų kovariacijų matrica pasidaro nebeįstrižaininė:

$$\Omega = E(uu') = \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\nu^2 (I_N \otimes I_T)$$

- Paklaidų variacija išlieka homoskedastinė, bet atsiranda serijinė koreliacija atskiriems individams:

$$\begin{aligned} \mathrm{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_\nu^2, & \text{kai } i = j,\ t = s \\ &= \sigma_\mu^2, & \text{kai } i = j,\ t \neq s \end{aligned}$$

- Reikia taikyti GLS, tam reikia Ω^{-1}.

Kovariacijų matricos spektrinis išdėstymas

Matricą Ω galima užrašyti taip sumą:

$$\Omega = \sigma_1^2 P + \sigma_\nu^2 Q,$$

čia $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$. Tada

$$\Omega^{-1} = \frac{1}{\sigma_1^2} P + \frac{1}{\sigma_\nu^2} Q$$
$$\Omega^{-1/2} = \frac{1}{\sigma_1} P + \frac{1}{\sigma_\nu} Q,$$

$y* = \sigma_\nu \Omega^{-1/2} y$ tipinis elementas yra $y_{it} - (1 - \sigma_\nu/\sigma_1)\bar{y}_{i.}$

AE modelio koeficientų įverčiai

- Padauginę regresijos modelį iš $\sigma_\nu \Omega^{-1/2}$, GLS gausime taikydami svorinį OLS.
- GLS sprendinys yra

$$\hat{\beta}_{GLS} = W_1 \tilde{\beta}_{Within} + W_2 \hat{\beta}_{Between},$$

  čia $W_1 + W_2 = I$.
- Matricos W_1 ir W_2 priklauso nuo $\phi = \sigma_\nu / \sigma_1$, taigi reikia rasti σ_1 ir σ_ν įverčius.
- Paprastai remiamasi tuo, kad $Pu \sim (0, \sigma_1^2 P)$ ir $Qu \sim (0, \sigma_\nu^2 Q)$.

AE modelio apžvalga

- Naudojamas, kai atsitikinai pasirenkame N individų.
- Kai $\sigma_\mu = 0$, GLS įvertis susiveda į OLS.
- Kai $T \to \infty$, $\hat{\beta}_{GLS}$ artėja į $\tilde{\beta}_{Within}$, $\mathrm{var}(\tilde{\beta}_{Within}) - \mathrm{var}(\hat{\beta}_{GLS})$ yra teigiamai apibrėžta matrica.
- Kai tikrasis modelis yra AE, GLS įverčiai su tikromis dispersijomis yra BLUE, o su įvertintomis yra asimptotiškai geriausi, kai $N \to \infty$ arba $T \to \infty$
- Nežinomom dispersijom surasti yra pasiūlyta daug metodų: Wallace ir Hussain(1969), Amemiya(1971), Swamy ir Arora (1972), Nerlove (1971). Metodai buvo palyginti Maddala ir Mount (1973), Taylor (1980). Buvo nustatyta, kad mažom imtims visi jie duoda panašius rezultatus.

Dviejų paklaidos komponentų modelis

- Įvedamas papildomas dėmuo:

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, N; \; t = 1, \ldots, T,$$

  čia μ_i – nestebimas individualus efektas, λ_t – nestebimas laiko efektas.

- Vektorinis pavidalas:

$$u = Z_\mu \mu + Z_\lambda \lambda + \nu,$$

  čia $Z_\mu = I_N \otimes \imath_T$, $Z_\lambda = \imath_N \otimes I_T$.

- Analogiškai nagrinėjami fiksuotų bei atsitiktinių efektų modeliai.

FE ir AE modeliai

- FE atveju įverčiai gaunami taikant LSDV, tik vidinės regresijos transformacijos matrica eliminuoja tiek individualius, tiek laiko efektus. Taip gauname įvertį $\tilde{\beta}$. Individualius ir laiko efektus gauname atitinkamai iš tarpinės individų ir tarpinės laiko regresijų.
- RE atveju matricą Ω galima panašiai išskaidyti į tarpusavyje ortogonalių idempotentinų matricų sumą. GLS β įvertis yra trijų regresijų svorinis vidurkis:

$$\hat{\beta}_{GLS} = W_1\beta_W + W_2\beta_B + W_3\beta_C$$

- Matricos W_1, W_2 ir W_3 taip pat priklauso nuo nežinomų dispersijų, kurias reikia įvertinti.

Testai duomenų panelškumui nustatyti

- Turime N regresijų

$$y_i = Z_i \delta_i + u_i$$

  Nulinė hipotezė $H_0 : \delta_i = \delta$ visiems i.

- Esant prielaidai $u \sim N(0, s^2 I_{NT})$ taikomas Chow testas:

$$F = \frac{(e'e - e^{*\prime} e^*)/(N-1)K}{e^{*\prime} e^*/N(T-K)},$$

  čia e - nulinės hipotezės regresijos liekanos, e^* - alternatyvios hipotezės regresijos liekanos

- Esant prielaidai $u \sim N(0, \Omega)$, taikomas Roy-Zellner testas, gaunamas Chow teste liekanas pakeitus GLS liekanų įverčiais.

- Klausimas ar traktuoti duomenis, kaip panelinius ar ne susiveda į hipotezės testavimą. Pritaikius apribojimus bus sumažinta variacija, bet gali atsirasti poslinkis, jeigu apribojimai neteisingi.
- Baltagi(1981) atliko Chow ir Roy-Zellner Monte-Carlo palyginimą. Esant dviejų komponentų paklaidos modeliui, Chow testas pernelyg dažnai atmetė nulinę hipotezę, kai ji buvo teisinga.
- Ziemer ir Weitzstein (1983) pasiūlė lyginti panelinių duomenų ($\hat{\delta}_{OLS}$) įverčius su nepaneliniais įverčiais ($\hat{\delta}_{i,OLS}$) pagal prognozavimo riziką

Testai individualiems ir laiko efektams nustatyti

- Breusch ir Pagan (1980) LM testas hipotezei $H_0 : \sigma^2_\mu = \sigma^2_\lambda = 0$

$$LM_1 = \frac{NT}{2(T-1)}\left[1 - \frac{\tilde{u}'(I_N \otimes J_T)\tilde{u}}{\tilde{u}'\tilde{u}}\right]^2$$

$$LM_2 = \frac{NT}{2(N-1)}\left[1 - \frac{\tilde{u}'(J_N \otimes I_T)\tilde{u}}{\tilde{u}'\tilde{u}}\right]^2,$$

  čia $\tilde{u}$ – OLS liekanos.

- $LM = LM_1 + LM_2$ esant teisingai H_0 yra pasiskirstęs pagal χ^2_2.
- Su LM_1 galima testuoti hipotezę $H_0 : \sigma^2_\mu = 0$, su LM_2 – $H_0 : \sigma^2_\lambda = 0$.

- Honda (1985) pasiūlė tolygiai galingiausius testus:

$$A = \sqrt{\frac{NT}{2(T-1)}}\left[\frac{\tilde{u}'(I_N \otimes J_T)\tilde{u}}{\tilde{u}'\tilde{u}} - 1\right]$$

$$B = \sqrt{\frac{NT}{2(N-1)}}\left[\frac{\tilde{u}'(J_N \otimes I_T)\tilde{u}}{\tilde{u}'\tilde{u}} - 1\right],$$

  tikrinti atitinkamoms nulinėms hipotezėms $H_0 : \sigma_\mu^2 = 0$ ir $H_0 : \sigma_\lambda^2 = 0$. A ir B yra asimptotiškai normalūs.
- Testai su vienpuse alternatyva hipotezei $H_0 : \sigma_\mu^2 = \sigma_\lambda^2 = 0$ tikrinti yra gaunami įvairiai sudėjus A ir B. Honda (1985), Baltagi, Chang ir Li (1992), Gourieroux, Holy, Monfort (1982).

Kiti testai

- Baltagi, Chang ir Li pasiūlė LM testus nulinėms hipotezėms $H_0 : \sigma_\mu^2 = 0$(tariant $\sigma_\lambda^2 > 0$), bei $H_0 : \sigma_\lambda^2 = 0$(tariant $\sigma_\mu^2 > 0$) tikrinti.
- Moulton ir Randolph (1989) parodė, kad FE modeliui naudojami F testai gali būti pritaikyti ir AE modeliui.
- Gourieroux, Holy, Monfort išvedė LR statistikas ir parodė, kad jos turi tą patį asimptotinį pasiskirstymą kaip ir LM statistikos.

Testų palyginimas

- Baltagi, Chang ir Li (1992) atliko įvairių testų Monte-Karlo palyginimą.
- Jei $H_0 : \sigma_\mu^2 = 0$ yra teisinga, bet $\sigma_\lambda^2 > 0$, visi testai į tai neatsižvelgiantys per dažnai atmeta nulinę hipotezę. Dideliems σ_μ^2 tai įtakos nedaro.
- Testai skirti $H_0 : \sigma_\mu^2 = 0$(tariant $\sigma_\lambda^2 > 0$) dideliems $\sigma_\mu^2 = 0$ turi didelę galią. Be to, net jei tikrajam modeliui $\sigma_\lambda^2 = 0$, testams tai įtakos nedaro.
- Palyginus su LM ir LR statistikomis F statistikos elgiasi šiek tiek geriau.

Hausmano specifikacijos testas

- Paklaidų komponentų modeliamas svarbi sąlyga yra $E(u_{it}|X_{it}) = 0$
- Jei ši sąlyga netenkinama, tai $\hat{\beta}_{GLS}$ yra paslinktas ir nesuderintas, bet $\tilde{\beta}_{Within}$ išlieka suderintas ir nepaslinktas.
- Hausman (1978) pasiūlė statistiką, besiremiančią šių įverčių skirtumu $\hat{q}_1 = \hat{\beta}_{GLS} - \tilde{\beta}_{Within}$:

$$m_1 = \hat{q}_1'[\mathrm{var}(\hat{q}_1)]\hat{q}_1,$$

esant nulinei hipotezei m_1 pasiskirstęs kaip χ^2_K.

Kiti specifikacijos testai

- Hausman ir Taylor (1981) parodė, kad nulinę hipotezę $E(u_{it}|X_{it}) = 0$ galima tikrinti naudojant bet kurį iš šių skirtumų: $\hat{q}_1 = \hat{\beta}_{GLS} - \tilde{\beta}_{Within}$, $\hat{q}_2 = \hat{\beta}_{GLS} - \tilde{\beta}_{Between}$ ir $\hat{q}_3 = \hat{\beta}_{Within} - \tilde{\beta}_{Between}$.
- Arellano (1993) sukonstravo testą, kuris nejautrus bet kokiai autokoreliacijos ir heteroskedastiškumo formai.
- Chamberlain (1982), Angrist ir Newey (1991) nagrinėjo specifikacijos testus FE modeliui. Chamberlain parodė, kad FE modelyje koeficientai iš tikrųjų yra apriboti.