# Ekonometrinė panelinių duomenų analizė

Vaidotas Zemlys

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

2005 m. balandžio 5 d.

# Individualieji efektai

- Mazodier ir Trognon (1978) apibendrino AE modelį, heteroskedastiškų individualių efektų atvejui $\mu_i \sim \text{IID}(0, w_i^2)$, su homoskedastinėmis paklaidomis $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$. Šiuo atveju atitinkamas kovariacijų matricos išskaidymas yra

$$\Omega = \text{diag}[Tw_i^2 + \sigma_\nu^2] \otimes \bar{J}_T + \text{diag}[\sigma_\nu^2] \otimes E_T$$

- Norint rasti GLS įverčius reikia įvertinti nežinomas dispersijas: $\sigma_\nu^2$ ir $w_i^2$, $i = 1, \dots, N$. Taigi reikia, kad $T >> N$.
- $\sigma_\nu^2$ galima įvertinti, kaip ir homoskedastiniu atveju, $w_i^2$ įverčius galima gauti naudojantis tuo, kad $\text{var}(u_{it}) = w_i^2 + \sigma_\nu^2$

# Paklaidų heteroskedastiškumas

- Analogiškai galima nagrinėti atvejį, kai paklaidos yra heteroskedastiškos paklaidos $\nu_{it} \sim \mathrm{IID}(0, w_i^2)$, o individualieji efektai homoskedastiniai – $\mu_i \sim \mathrm{IID}(0, \sigma_\mu^2)$. Tada kovariacijų matricos išskaidymas yra

$$\Omega = \mathrm{diag}[\sigma_\mu^2] \otimes \bar{J}_T + \mathrm{diag}[w_i^2] \otimes E_T$$

- Randolph (1988) nagrinėjo bendresnius AE modelius, kai tiek paklaidos tiek individualūs efektai yra heteroskedastiški.

# AR(1) paklaidų modelis

- Klasikiniam AE modelyje laike koreliuoja tik to pačio individo stebėjimai, ir ši koreliacija yra pastovi visiems laiko tarpams.
- Lillard ir Willis (1978) apibendrino AE modelį ($\mu_i \sim \mathrm{IID}(0, \sigma_\mu^2)$) serijinės koreliacijos atvejui, tardami, kad $\nu_{it}$ yra AR(1) laiko eilutė:

$$\nu_{it} = \rho \nu_{i,t-1} + \epsilon_{it},$$

čia $|\rho| < 1$ ir $\epsilon_{it} \sim IID(0, \sigma_\epsilon^2)$. Individualūs efektai $\mu_i \sim \mathrm{IID}(0, \sigma_\mu^2)$ nepriklauso nuo $\nu_{it}$, ir $\nu_{i0} \sim \mathrm{IID}(0, \sigma_\epsilon^2/(1-\rho^2))$

# Koeficientų vertinimas AR(1) atveju

- Naudodami Prais-Wainsten transformaciją $C$ pakeičiame AR(1) paklaidas į paklaidas be serijinės koreliacijos. Šią transformaciją reikia pritaikyti $N$ individams. Modelio paklaidos tada yra

$$u^* = (I_N \otimes C\imath_T)\mu + (I_N \otimes C)\nu$$

- Šitų paklaidų kovariacijų matricai $\Omega*$ egzistuoja atitinkamas išskaidymas, kurio pagalba lengvai galime rasti $\Omega*^{-1/2}$. Turėdami $\Omega*^{-1/2}$ galime gauti GLS įverčius.
- AR(1) atveju matrica $\Omega*^{-1/2}$ priklausys nuo $\sigma_\mu^2$, $\sigma_\epsilon^2$ ir $\rho$, taigi norint gauti GLS įverčius reikia gauti šių dydžių įverčius.

# Kiti laiko eilučių paklaidų modeliai

- Tegu $\exists C$ tokia, kad $\mathrm{var}((I_N \otimes C)\nu) = \sigma^2 I_N T$. Modelį $y_{it} = X'_{it}\beta + \mu_i + \nu_{it}$ transformuokime tokia dviejų žingsnių transformacija:
    1. $y_i^* = Cy_i$, $y_i' = (y_{i1}, \ldots, y_{iT})$. (Eliminuojama serijinė koreliacija).
    2. $$y_{it}^{**} = y_{it}^* - \theta_\alpha \alpha_t \left( \sum_{i=1}^{T} \alpha_s y_{is}^* \right) / \left( \sum_{i=1}^{T} \alpha_s^2 \right),$$

       čia $C\imath_T = (\alpha_1, \ldots \alpha_T)$, $\theta_\alpha = 1 - (\sigma/\sigma_\alpha)$ ir $\sigma_\alpha^2 = \sigma_\mu^2 (\sum_{i=1}^{T} \alpha_T) + \sigma$.

  OLS įverčiai taip transformuotam modeliui duos GLS įverčius pradiniam modeliui.
- Šį metodas pritaikomas AR(2), MA(1) paklaidų modeliams. Baltagi ir Li (1994) apibendrino atvejui MA(q), Galbraith ir Zinde-Walsh (1995) atvejui ARMA(p,q).

# Testas serijinei koreliacijai ir individualiems efektams nustatyti

- Baltagi ir Li (1995) LM testas jungtinei hipotezei $H_0 : \sigma_\mu^2 = 0, \rho = 0$ tikrinti:

$$LM = \frac{NT^2}{2(T-1)(T-2)}[A^2 - 4AB + 2TB^2],$$

čia $A = [\widehat{u}'(I_N \otimes J_T)\widehat{u}/(\widehat{u}'\widehat{u})] - 1$, $B = \widehat{u}'\widehat{u}_{-1}/\widehat{u}'\widehat{u}$. Dideliems $N$ šios statistikos pasiskirstymas yra $\chi_2^2$, jei teisinga nulinė hipotezė.

- Hipotezei $H_0 : \sigma_\mu^2 = 0, \lambda = 0$ tikrinti tinka ta pati statistika.
- Naudojant $A$ ir $B$ galime tikrinti hipotezes $H_0 : \sigma_\mu^2 = 0$ ir $H_0 : \rho = 0$ (arba $\lambda = 0$).

# Testas serijinei koreliacijai, AE ir FE modeliuose

- Baltagi ir Li (1995) išvedė testą hipotezei $H_0 : \rho = 0$ (kai $\sigma_\mu^2 > 0$) tikrinti. Kadangi tariama, kad yra AE modelis, reikia naudoti GLS įverčius ir LM statistika pasidaro sudėtinga.
- Testas hipotezei $H_0 : \rho = 0$, kai $\mu_i$ – fiksuoti:

$$LM = \sqrt{NT^2/(T-1)}(\widetilde{\nu}'\widetilde{\nu}_{-1}/\widetilde{\nu}'\widetilde{\nu})$$

  čia $\widetilde{\nu}$ – vidinės regresijos liekanos ir statistika dideliems $T$ pasiskirsčius kaip $N(0,1)$.
- Abiem atvejais statistika ta pati, nepriklausomai nuo to, kokį serijinės koreliacijos modelį testuosime, ar $AR(1)$ ar $MA(1)$.

# Durbin-Watson statistika FE modeliui

- Bhargava, Franzini ir Narendranathan (1982) pasiūlė testą hipotezei $H_0 : \rho = 0$ su alternatyva $|\rho| < 1$:

$$d_p = \sum_{i=1}^{N} \sum_{t=2}^{T} (\widetilde{\nu}_{it} - \widetilde{\nu}_{i,t-1})^2 / \sum_{i=1}^{N} \sum_{t=1}^{T} (\widetilde{\nu}_{it}^2,$$

  ir parodė, kad jis yra lokaliai galingiausias invariantiškas testas $\rho = 0$ aplinkoje, nepriklausomai nuo pasirinktų regresorių.
- Tie patys autoriai pasiūlė lokaliai galingiausią testą $\rho = 1$ aplinkoje, bei testą hipotezei $H_0 : \rho = 1$ tikrinti. Autoriai parodė, kad dideliems $N$ visi šie trys testai nagrinėjami esant nulinei atsitiktinio klaidžiojamo hipotezei sutampa.

# Testai serijinės koreliacijos pobūdžiui nustatyti

- Baltagi ir Li (1995) pasiūlė du apibendrinimus Burke, Godfrey ir Termayne (1990) testui atskirti $AR(1)$ nuo $MA(1)$ proceso.
- Esant nulinei $AR(1)$ hipotezei $\mathrm{corr}(\nu_{it}, \nu_{i,t-\tau}) = \rho_\tau = \rho_1^\tau$. Taigi $AR(1)$ procesui $\rho_2 - (\rho_1)^2 = 0$. $MA(1)$ procesui $\rho_2 = 0$ ir $\rho_2 - (\rho_1)^2 < 0$.
- $\rho_s$, $s = 1, 2$ iverčius galime gauti kiekvienam individui naudojant vidinės regresijos liekanas:

$$(\widetilde{\rho}_s)_i = \sum_{t=2}^{T} \widetilde{u}_{it}\widetilde{u}_{i,t-s} / \sum_{t=2}^{T} \widetilde{u}_{it}^2.$$

Taikant tokius iverčius gauname $O(1/T)$ poslinkį, todėl šia idėja besiremianti statistika netiks tradiciniams panelinių duomenų modeliams, kuriems $T$ mažas, o $N$ - didelis.

# Kitas testas

- Nulinė hipotezė $H_0 : \nu_{it} = \epsilon_{it} + \lambda\epsilon_{i,t-1}$
- Tegu $Q_s = \sum\sum u_{it}u_{i,t-s}/N(T-s)$. Kai $N \to \infty$
  1. $MA(1)$ modeliui

$$\begin{aligned}\text{plim } Q_0 &= \sigma_\mu^2 + \sigma_\epsilon^2(1+\lambda^2)\\ \text{plim } Q_1 &= \sigma_\mu^2 + \lambda\sigma_\epsilon^2\\ \text{plim } Q_s &= \sigma_\mu^2, \text{ kai } s \geq 2.\end{aligned}$$

  2. $AR(1)$ modeliui

$$\begin{aligned}\text{plim } Q_0 &= \sigma_\mu^2 + \sigma_\nu^2\\ \text{plim } Q_1 &= \sigma_\mu^2 + \rho^s\sigma_\nu^2, \text{ kai } s \geq 1.\end{aligned}$$

Dideliems $N$, $\sqrt{N/V}(Q_2 - Q_3)$ turi normalų pasiskirstymą, čia $V$ – tam tikras normuojantis dydis.

# SUR su paklaidų komponentų modeliu

- Avery (1977) nagrinėjo $M$ lygčių $y_j = Z_j\delta_j + u_j$ modelį su paklaidomis $u_j = Z_\mu \mu_j + \nu_j$. Lygčių tarpusavio priklausomybė išreiškiama taip:

$$E\begin{pmatrix}\mu_j \\ \nu_j\end{pmatrix}(\mu_j', \nu_j') = \begin{bmatrix}\sigma^2_{\mu_{jl}} I_N & 0 \\ 0 & \sigma^2_{\nu_{jl}} I_{NT}\end{bmatrix}$$

- Visų $M$ lygčių kovariacijų matrica tada yra

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

- Toliau vertiname panašiai kaip ir vienos regresijos atveju. Jei visoms lygtims regresorių matrica vienoda, tai visos sistemos įvertis nesutampa su atskirų lygčių iverčiais kaip modeliui be paklaidos komponentų.