Nestacionarūs panelinių duomenų modeliai

Vaidotas Zemlys

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

2005 m. gegužės 3 d.

Panelinių duomenų statistikų asimptotika

- Dažniausiai panelinių duomenų analizėje sutinkamas procesas yra

$$X_{n,T} = \frac{1}{k_n} \sum_{i=1}^{n} Y_{i,T},$$

čia $Y_{i,T}$ yra i atžvilgiu nepriklausomi atsitiktiniai vektoriai, ir paprastai, tai standartizuota panelinių duomenų laiko eilučių dalies suma.

- Jeigu turime paprastą panelinės regresijos modelį su individualias efektais:

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}$$

tai dažniausiai nagrinėjamas vidinės regresijos OLS β įvertis:

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(u_{it} - \bar{u}_{i.})(x_{it} - \bar{x}_{i.})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i.})^2}$$

Įvairūs konvergavimai

- Sekvencinis konvergavimas. Tegu egzistuoja $Y_{i,T}$ ribos Y_i. Imant $T \to \infty$ randamos tarpinės ribos $X_n = \frac{1}{k_n} Y_i$. Tada imant $n \to \infty$ randama X_n riba. Gautas ribinis atsitiktinis dydis vadinamas sekvencine riba.
- Konvergavimas pagal diagonalines trajektorijas. Laikoma, kad $T = T(n)$ ir ieškoma ribos, kai $n \to \infty$.
- Bendras konvergavimas. Netaikomi jokie apribojimai n ir T, ir ieškoma riba, kai n ir T į begalybę artėja kartu.

- Tegu $Z_{i,t} = N(0, 1\chi\{i < t\} + i\chi\{i \geq t \geq 1\})$, $Y_{i,T} = \frac{1}{\sqrt{T}} Z_{i,t}$ ir $X_{n,T} = \frac{1}{\sqrt{n}}$. Tada sekvencinė $X_{n,T}$ riba yra $N(0,1)$, bet

$$X_{T^r,T} \Rightarrow \begin{cases} N(0,1), & \text{kai } r < 1/2, \\ N(0,4/3), & \text{kai } r = 1/2, \\ \text{nekonverguoja}, & \text{kai } r > 1/2, \end{cases}$$

kai $T \to \infty$.
- Tegu $X_{n,T} \Rightarrow X_n$, kai $T \to \infty$ ir $X_n \Rightarrow X$, kai $n \to \infty$, tada $X_{n,T} \Rightarrow X$, kai $T, n \to \infty$, tada ir tik tada, kai

$$\limsup_{n,T} |E(f(X_{n,T}) - f(X))|, \ \forall f \in \mathcal{C}$$

Levin-Lin testai

- Nagrinėjo modelį

$$y_{it} = \rho_i y_{i,t-1} + z_{it}'\gamma + u_{it}$$

- Tegu $z_t = (z_{1t}, \ldots, z_{Nt})'$, $h(t,s) = z_t' \left(\sum_{t=1}^{T} z_t z_t' \right) z_s$, $\tilde{u}_{it} = u_{it} - \sum_{s=1}^{T} h(t,s) u_{is}$ ir $\tilde{y}_{it} = y_{it} - \sum_{s=1}^{T} h(t,s) y_{is}$. Tada

$$\sqrt{N}T(\hat{\rho} - 1) = \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \tilde{y}_{i,t-1} \tilde{u}_{it}}{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{y}_{i,t-1}^2}$$

$$t_\rho = \frac{(\hat{\rho} - 1)\sqrt{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{i,t-1}^2}}{s_e},$$

čia $s_e^2 = (1/NT) \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{u}_{it}^2$

Levin ir Lin testų ribiniai pasiskirstymai

- Esant nulinei hipotezei $H_0 : \rho = 1$
  Levin ir Lin gavo tokius ribinius pasiskirstymus:

z_{it}	$\hat{\rho}$	$f(t_\rho) \Rightarrow N(0,1)$
0	$\sqrt{N}T(\hat{\rho}-1) \Rightarrow N(0,2)$	t_ρ
1	$\sqrt{N}T(\hat{\rho}-1) \Rightarrow N(0,2)$	t_ρ
μ_i	$\sqrt{N}T(\hat{\rho}-1) + 3\sqrt{N} \Rightarrow N(0, \frac{51}{5})$	$\sqrt{1.25}t_\rho + \sqrt{1.875N}$
$(\mu_i, t)'$	$\sqrt{N}(T(\hat{\rho}-1) + 7.5) \Rightarrow N(0, \frac{2895}{112})$	$\sqrt{\frac{488}{277}}(t_\rho + \sqrt{3.75N})$

Harris ir Tzavalis (1999)

- Nagrinėjo atvejį, kai T - fiksuotas. Tada

z_{it}	$\hat{\rho}$
0	$\sqrt{N}(\hat{\rho}-1) \Rightarrow N(0, \frac{2}{T(T-1)})$
μ_i	$\sqrt{N}\left(\hat{\rho}-1+\frac{3}{T+1}\right) \Rightarrow N(0, \frac{3(17T^2-20T+17)}{5(T-1)(T+1)^3}\Big)$
$(\mu_i, t)'$	$\sqrt{N}\left(\hat{\rho}-1+\frac{15}{2(T+2)}\right) \Rightarrow N(0, \frac{15(193T^2-728T+1147)}{112(T+2)^3(T-2)}\Big)$

Im, Pesaran ir Shin (1997) testai

- Modelis

$$y_{it} = \rho_i y_{i,t-1} + \sum_{j=1}^{p_i} \varphi_{ij} \Delta y_{i,t-j} + z_{it}' \gamma + \varepsilon_{it}$$

- Pasiūlė suvidurkinti ADF statistikas:

$$\bar{t} = \frac{1}{N} \sum_{t=1}^{N} t_{\rho_i}$$

- Kadangi $t_{\rho_i} \Rightarrow t_{it}$, kai $T \to \infty$, tai centruotas $\bar{t} \Rightarrow N(0,1)$ pagal CRT, tarus, kad t_{it} yra IID.

LL ir IPS testų palyginimas

- LL testas reikalauja, kad ρ būtų vienodas visoms šalims
- Abu reikalauja $N/T \to 0$. Pažeidus šią sąlygą testų dydis išsikreipia, kai N didelis palyginus su T.
- Breitung (2000) nagrinėjo LL ir IPS testų lokalią galią, esant lokalių alternatyvų sekai. Jeigu įtraukiami individualūs trendai, tai LL ir IPS testų galia stipriai sumažėja. Monte-Karlo rezultatai rodo, kad LL ir IPS testai labai jautrūs z_{it} parinkimui.

Testų p-reikšmių jungimas

- Tegu G_{iT_i} yra vienetinės šaknies statistika i-tajai grupei, ir $G_{iT_i} \Rightarrow G_i$. Tegu $p_i = F_{G_i}(G_{iT_i})$. Maddala, Wu ir Choi (1999a) pasiūlė naudoti statistiką

$$P = -2 \sum_{i=1}^{N} \ln p_i$$

- Fiksuotiems N, P yra pasiskirsčius, kaip χ^2 su $2N$ laisvės laipsnių, kai $T_i \to \infty$. Šios statistikos trūkumas yra tas, kad jos p-reikšmes reikia skaičiuoti su Monte-Karlo simuliacijomis.
- Choi (1999) pasiūlė dar dvi panašias statistikas $Z = (1/\sqrt{N}) \sum_{i=1}^{N} \Phi^{-1}(p_i)$ ir $L = \sum_{i=1}^{N} \ln(p_i/(1-p_i))$.

Hadri LM testas

- Modelis

$$y_{it} = z_{it}'\gamma + e_{it},$$

čia $e_{it} = \sum_{j=1}^{t} u_i j + \varepsilon_{it}$.

- Konstruojama statistika:

$$LM = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T} S_{it}^2}{\hat{\sigma}_e^2}$$

- Esant nulinei hipotezei, kad modelis yra stacionarus:

$$LM \xrightarrow{p} E\left[\int W^2\right],$$

kai $T \to \infty$ ir po to $N \to \infty$.

Moon ir Philips (2000) modelis

- Nagrinėjo nestacionarų modelį

$$\begin{aligned} y_{it} &= \alpha_{i0} + \alpha_{i1} t + y_{it}^0 \\ y_{it}^0 &= \beta y_{i,t-1}^0 + u_{it}, \end{aligned}$$

  čia $\beta = \exp(c/T)$.
- Informacija apie c naudinga analizuojant vienetinės šaknies, kointegravimo testų galią.
- Parodė, kad kai $c \leq 0$, naudojant panelinius duomenis galima gauti suderintą c įvertį.

Ilgo nuotolio vidutiniai sąryšiai

- Jei (Y, X) yra dvimatis normalusis atsitiktinis dydis $N(0, \Sigma)$ su

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{xx} \end{bmatrix}$$

  tai Y nuo X regresijos koeficientas $\beta = \Sigma_{yx}\Sigma_{xx}^{-1}$.
- Klasikiniam regresiniam modelyje

$$Y_t = \beta X_t + U_t,$$

  čia $EX_t = EU_t = 0$ ir X_t su U_t - nekoreliuoti,
  $\beta = (EY_t X_t')(EX_t X_t') = \Sigma_{yx}\Sigma_{xx}^{-1}$

- Tegu $Z_t = (Y_t', X_t')'$ ir $Z_t = Z_{t-1} + U_t$. Tegu egzistuoja ilgo nuotolio kovariacijų matrica

$$\Omega = \lim_T E\left(\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T} U_t\right)\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T} U_t'\right)\right)$$

padalinta šitaip:

$$\Omega = \begin{bmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{yx} & \Omega_{xx} \end{bmatrix}$$

- Tada

$$\beta = \lim_T E\left(\frac{Y_T}{\sqrt{T}}\frac{X_T'}{\sqrt{T}}\right)\left[E\left(\frac{X_T}{\sqrt{T}}\frac{X_T'}{\sqrt{T}}\right)\right]^{-1} = \Omega_{yx}\Omega_{xx}^{-1}$$

nusako ilgo nuotolio sąryšį tarp nestacionarių kintamųjų X_t ir Y_t.

Ilgo nuotolio sąryšiai paneliniams duomenims

- Laikome, kad turime imtį $Z_{i,t}$, tokią, kad $\Omega = E\Omega_i$.
- Philips ir Moon (1999) nagrinėjo vertino β šiais atvejais
  1. Ω_i yra teigiamai apibrėžta kiekvienam i.
  2. Ω_i yra nepilno rango ir kiekvienam i kointegracija skirtinga.
  3. Ω_i yra nepilno rango ir kiekvienam i kointegracija vienoda.

  Visiems šiems atvejams OLS β įvertis yra suderintas, kai $n, T \to \infty$, bet norint gauti ribinį pasiskirstymą reikia prielaidos $n/T \to \infty$.

Kao (1999) testai

- Nagrinėja modelį

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + e_{it}$$

  čia $x_{it} = x_{i,t-1} + \varepsilon_{it}$ ir e_{it} – $I(1)$. Konstruoja DF testus iš regresijos liekanų:

$$\hat{e}_{it} = \rho\hat{e}_{i,t-1} + \nu_{it}$$

  Tada kointegruotumo nebuvimas išreiškiamas nuline hipoteze $H_0 : \rho = 1$.

- Pasiūlė 4 testus remiantis OLS įverčiu $\hat{\rho}$ ir t_ρ, kai $z_{it} = \mu_i$, esant stipriai egzogeniškiems arba endogeniniams regresoriams.

McCoskey ir Kao (1998) testas

- Modelis

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it}'\beta + e_{it} \\
x_{it} &= x_{i,t-1} + \varepsilon_{it} \\
e_{it} &= \gamma_{it} + u_{it} \\
\gamma_{it} &= \gamma_{i,t-1} + \theta u_{it}
\end{aligned}
$$

- Nulinė kointegruotumo hipotezė yra ekvivalenti $\theta = 0$. Testas apibrėžiamas taip:

$$LM = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T} S_{it}^2}{\hat{\sigma}_e^2}$$

čia $S_{it} = \sum_{j=1}^{t}\hat{e}_{it}$.
- Ribinis statistikos pasiskirstymas yra

$$\sqrt{N}(LM - \mu_\nu) \Rightarrow N(0, \sigma_\nu^2),$$

čia μ_ν ir σ_ν^2 yra Brauno judesio funkcionalai.

Kiti testai

- Kaip ir vienetinės šaknies testų atveju, galima nagrinėti kointegravimo testų p-reikšmių statistikas.
- Larson, Lyhagen, Löthgren (2001) nagrinėjo LR testus kointegravimo rangui nustatyti.
- Groen ir Kleibergen (1999) nagrinėjo panelinių duomenų kointegruotumą pastoviam skaičiui VEC (vector error correction) modelių.
- Hall, Lazarova ir Urga (1999) taikė principinių komponentų analizę vienodų stochastinių trendų skaičiui nustatyti.